A Publication of the American Psychological Association

# Journal of Abnormal and Social Psychology

Founded by Morton Prince, 1906

CONTENTS

Articles:

The Open System in Personality Theory ..... *Gordon W. Allport*

On the Reliability of the Anamnestic Interview ..... *Bruce A. Bradford, Anne Brodsky, and Ann Jones Shaw*

Acquiescent Response Set and the Psychoanalytic Approach to Analogy Experiments ..... *James E. Keeler*

Affect-Stimulating Effects of Caffeine ..... *Robert J. Prentice*

Authoritarianism and Ego Resistance to Sexuality and Closeness to Mother ..... *Ralph Reichard*

Group Interaction and Perceived Similarity of Members ..... *Jane Stafford, Nicholas R. Cotton, and Gerald Gordon*

Verbal Rate, Eyeblink, and Content in Psychiatric Interviews ..... *Frederick H. Kanfer*

The Susceptibility of Chronic Schizophrenics to Social Influence in the Formation of Perceptual Judgments ..... *Carmi Schooler and Robert E. Spohn*

Differences in Response Latency and Response Variability Between Repressors and Low Anxious Subjects in a Flicker-Fusion Task ..... *Robert A. Wyer, Jr.*

Leader-Member Interactions in Management Consultation ..... *Lawrence Solomon, Jay R. Jackson, and Jack Morrison*

Departments are inside front cover

This is the last issue of Volume 61.

Next Page [illegible] begins the new Volume with Number 1.

Volume 61 · NOVEMBER 1960 · Number 3

CONTENTS—*Continued*


Competence and Conformity in the Acceptance of Influence ..... 365
*E. P. Hollander*

Creativity and Primary Process: A Study of Adaptive Regression ..... 370
*Frank Barron and Robert R. Holt*

Effects of Stress on Goal Setting Behavior ..... 390
*Kenneth L. Rau and Roger G. Russell*

Rigidity in Neurosis and General Personality Functioning ..... 399
*Lawrence A. Pervin*

Two Correlates of Symbolic Sexual Arousal: Level of Society and Liking for Esthetic Materials ..... 404
*Michael A. Wallach*

Changes in Interpersonal Perception as a Means of Resolving Cognitive Dissonance ..... 409
*Keith E. Davis and Edward E. Jones*

Some Personal and Situational Determinants of Yielding to Influence ..... 411
*Archibald S. Patel and Irene E. Carlson*

The Attribution of Similarity: The Influence of Success and Failure ..... 419
*Anthony J. Smith*

The Indirect Representation of Sexual Stimuli in Schizophrenic and Normal Subjects ..... 424
*Leo B. Arey*

The Unconscious Perception of the Meaning of Verbal Stimuli ..... 430
*Marcus J. Poser and Charles W. Eriksen*

The Child's Differential Perception of Parental Attributes ..... 440
*Jerome Kagan and Judith Lemkin*

Effects of White Anti-Contrarianism in Biracial Work Groups ..... 448
*Irwin Katz and Lawrence Benjamin*

Size Estimation of Pictures Associated with Success and Failure as a Function of Manifest Anxiety ..... 457
*Theodore P. Zahn*

The Effect of Aroused Motivation on Attitudes ..... 466
*Helen Peak*

Audience Stress, Personality, and Speech ..... 469
*Harry Lenn, Alfred L. Baldwin, Mary Gallway, and Allan Pointe*


CASE REPORTS


The Contextual Associates of Sun and God in Schreber's Autobiography ..... 474
*Julius Laffal*

A Clinical Note on the Specific Nature of an "Only" Trauma ..... 480
*Bertram P. Karon*


CRITIQUES AND NOTES


Male and Female Subjects in Personality Research ..... 482
*Carlton and Rae Carlson*

Personality Similarity, Positive Affect, and Interpersonal Attraction ..... 484
*Carroll E. Izard*

The Stability of Cognitive Controls ..... 485
*Riley W. Gardner and Robert T. Long*

Prestige Effects in Body-Sway Suggestibility ..... 488
*J. C. Orr*

Social Conformity in Perception and the Autokinetic Effect ..... 489
*Norman S. Endler*

A Factor Analysis of the Wesley Rigidity Inventory: Its Relationship to Age and Nonverbal Intelligence ..... 491
*Judith M. Chown*

Intervention-Intervention and Kinesthetic Aftereffects ..... 493
*Alan Reichenbach and Lawrence J. Bootzin*

Volume 61 November 1960 Number 3

# Journal of Abnormal and Social Psychology

Editor: M. Brewster Smith, *University of California, Berkeley*
Associate Editor: E. J. Shoben, Jr., *Teachers College, Columbia University*

Consulting Editors

Gordon W. Allport
*Harvard University*

Urie Bronfenbrenner
*Cornell University*

Roger Brown
*Massachusetts Institute of Technology*

Donald T. Campbell
*Northwestern University*

Launor F. Carter
*System Development Corporation*

Richard Christie
*Columbia University*

Morton Deutsch
*Bell Telephone Laboratories*

I. E. Farber
*State University of Iowa*

Donald W. Fiske
*University of Chicago*

N. L. Gage
*University of Illinois*

Roger W. Heyns
*University of Michigan*

Robert R. Holt
*New York University*

Irving L. Janis
*Yale University*

Harold H. Kelley
*University of Minnesota*

Leonard S. Kogan
*Community Service Society of New York*

Richard S. Lazarus
*University of California, Berkeley*

Daniel J. Levinson
*Harvard Medical School*

David C. McClelland
*Harvard University*

Leo Postman
*University of California, Berkeley*

Eliot H. Rodnick
*Duke University*

Nevitt Sanford
*University of California, Berkeley*

Seymour Sarason
*Yale University*

Richard L. Solomon
*University of Pennsylvania*

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis, those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued bimonthly in January, March, May, July, September, and November, two volumes per year. The two volumes average 900 pages. The subscription price per year is $16.00; foreign, $16.50. Single issues are $3.00.

Original manuscripts should be sent to the Editor-elect:

DANIEL KATZ
Center for Advanced Study in
the Behavioral Sciences
202 Junipero Serra Boulevard
Stanford, California

---

ARTHUR C. HOFFMAN
Managing Editor

HELEN ORR
Promotion Manager

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street, N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Mt. Royal and Guilford Aves., Baltimore 2, Md.
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Second-class postage paid at Baltimore, Maryland, and at additional mailing offices

Copyright 1961 by the American Psychological Association, Inc.

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 301-310

# THE OPEN SYSTEM IN PERSONALITY THEORY[1]

GORDON W. ALLPORT
*Harvard University*

OUR orb of common interest has two faces—one turned toward social psychology, the other toward personality. As things stand at the moment the first visage seems to me slightly depressed; the second slightly manic. However that may be, I should like to start this discussion by bringing the two faces into confrontation.

My first observation is that as members of Division 8 we are conspicuously the victims of fashion. Although our persona is sedate and seemly, we have our own hula hoops, flying saucers, and our own way of flagpole sitting. The interquartile range of our crazes I estimate to be about 10 years. McDougall's instinct theory held sway from 1908 to approximately 1920. Watsonian behaviorism dominated the scene for the next decade; then habit hierarchies, then field theory; now phenomenology. We never seem to solve our problems or exhaust our concepts; we only grow tired of them. At the moment it is fashionable to investigate response-set, coding, sensory deprivation, person perception, and to talk in terms of system theory—a topic to which we shall soon return. Ten years ago fashion called for group dynamics, Guttman scales, and for research on the centipedal properties of the authoritarian personality. Twenty years ago it was frustration-aggression, Thurstone scales, and national morale. And so it goes. We are fortunate that each surge of fashion leaves a rich residue of gain.

Fashions have their amusing and their serious side. We can smile at the way bearded problems receive tonsorial transformation. Having tired of "suggestibility" we adopt the new hairdo known as "persuasibility." Modern ethology excites us, and we are not troubled by the recollection that a century ago John Stuart Mill staked down the term to designate the new science of human character. We like the neurological concept of "gating," conveniently forgetting that American functionalism always stood firm for the dominance of general mental sets over specific. Reinforcement appeals to us but not the age-long debate over hedonism. The problem of freedom we brush aside in favor of "choice points." We avoid the body-mind problem but are in fashion when we talk about "brain models." Old wine, we find, tastes better from new bottles.

The serious side of the matter enters when we and our students forget that the wine is indeed old. Picking up a recent number of the *Journal of Abnormal and Social Psychology* I discover that the 21 articles written by American psychologists confine 90% of their references to publications of the past 10 years, although most of the problems they investigate have gray beards. In the same issue of the journal, three European authors locate 50% of their references prior to 1949. What this proves I do not know, excepting that European authors were not born yesterday. Is it any wonder that our graduate students reading our journals conclude that literature more than a decade old has no merit and can be safely disregarded? At a recent doctoral examination the candidate was asked what his thesis on physiological and psychological conditions of stress had to do with the body-mind problem. He confessed he had never heard of the problem. An undergraduate said that all he knew about Thomas Hobbes was that he sank with the *Leviathan* when it hit an iceberg in 1912.

## A PSYCHOLINGUISTIC TRIFLE

Our windows are pretty much closed toward the past, but we rightly rejoice in our growth since World War II. Among the many fortunate developments is rejuvenation in the field of psycholinguistics. (Even here, however, I cannot refrain from pointing out that the much discussed Whorfian hypothesis was old stuff in the days of Wundt, Jespersen, and Sapir.) Be that as it may, I shall introduce my discussion of open systems in personality theory by a crude Whorfian analysis of our

[1] Invited address to Division 8—Personality and Social Psychology of the American Psychological Association, September 4, 1959.

own vocabulary. My research (aided by the kind assistance of Stanley Plog) is too cursory to warrant detailed report.

What we did, in brief, was to study the frequency of the prefix *re-* and of the prefix *pro-* in psychological language. Our hypothesis was that *re-* compounds, connoting as they do again-ness, passivity, being pushed or maneuvered, would be far more common than *pro-* compounds connoting futurity, intention, forward thrust. Our sample consisted of the indexes of the *Psychological Abstracts* at 5-year intervals over the past 30 years; also all terms employing these prefixes in the Hinsie and Shatzky *Psychiatric Dictionary* and in the English and English *Psychological Dictionary*. In addition we made a random sampling of pages in five current psychological journals. Combining these sources it turns out that *re-* compounds are nearly five times as numerous as *pro-* compounds.

But, of course, not every compound is relevant to our purpose. Terms like reference, relationship, reticular, report do not have the connotation we seek; nor do terms like probability, process, and propaganda. Our point is more clearly seen when we note that the term reaction or reactive occurs hundreds of times, while the term proaction or proactive occurs only once—and that in English's *Dictionary*, in spite of the fact that Harry Murray has made an effort to introduce the word into psychological usage.

But even if we attempt a more strict coding of this lexical material, accepting only those terms that clearly imply reaction and response on one side and proaction or the progressive programing of behavior on the other, we find the ratio still is approximately 5:1. In other words our vocabulary is five times richer in terms like reaction, response, reinforcement, reflex, respondent, retroaction, recognition, regression, repression, reminiscence than in terms like production, proceeding, proficiency, problem solving, propriate, and programing. So much for the number of different words available. The disproportion is more striking when we note that the four terms reflex, reaction, response, and retention together are used 100 times more frequently than any single *pro-* compound excepting problem solving and projective—and this latter term, I submit, is ordinarily used only in the sense of reactivity.

The weakness of the study is evident. Not all terms connoting spontaneous, future oriented behavior begin with *pro-*. One thinks of expectancy, intention, purpose. But neither do all terms connoting passive responding or backward reference in time begin with *re-*. One thinks of coding, traces, input-output, and the like. But while our analysis leaves much to be desired it prepares the way for our critique of personality theory in terms of systems. The connecting link is the question whether we have the verbal, and therefore the conceptual, tools to build a science of change, growth, futurity, and potential; or whether our available technical lexicon tends to tie us to a science of response, reaction, and regression. Our vocabulary points to personality development from the past up to now more readily than to its development from here-on-out into the future.

## THE CONCEPT OF SYSTEM

Until a generation or so ago science, including psychology, was preoccupied with what might be called "disorganized complexity." Natural scientists explored this fragment and that fragment of nature; psychologists explored this fragment and that fragment of experience and behavior. The problem of interrelatedness, though recognized, was not made a topic for direct inquiry.

What is called system theory today—at least in psychology—is the outgrowth of the relatively new organismic conception reflected in the work of von Bertalanffy, Goldstein, and in certain aspects of gestalt psychology. It opposes simple reaction theories where a virtual automaton is seen to respond discretely to stimuli as though they were pennies-in-the-slot. Interest in system theory is increasing in psychology, though perhaps not so fast as in other sciences.

Now a system—any system—is defined merely as a complex of elements in mutual interaction. Bridgman (1959), as might be expected of an operationist, includes a hint of method in his definition. He writes, a system is "an isolated enclosure in which all measurements that can be made of what goes on in the system are in some way correlated" (p. 188).

Systems may be classified as closed or open. A closed system is defined as one that admits no matter from outside itself and is therefore subject to entropy according to the second law of thermodynamics. While some outside energies, such as change in temperature and wind may play upon a closed system, it has no restorative properties and no transactions with its environment, so that like a decaying bridge it sinks into thermodynamic equilibrium.

Now some authors, such as von Bertalanffy (1952b), Brunswik (1955), and Pumpian-Mindlin (1959), have said or implied that certain theories of psychology and of personality operate with the conception of closed systems. But in my opinion these critics press their point too far. We had better leave closed systems to the realm of physics where they belong (although even here it is a question whether Einstein's formula for the release of matter into energy does not finally demonstrate the futility of positing a closed cgs system even in physics). In any event it is best to admit that all living organisms partake of the character of open systems. I doubt that we shall find any advocate of a truly closed system in the whole range of personality theory. At the same time current theories do differ widely in the amount of openness they ascribe to the personality system.

If we comb definitions of open systems we can piece together four criteria: there is intake and output of both matter and energy; there is the achievement and maintenance of steady (homeostatic) states, so that the intrusion of outer energy will not seriously disrupt internal form and order; there is generally an increase of order over time, owing to an increase in complexity and differentiation of parts; finally, at least at the human level, there is more than mere intake and output of matter and energy: there is extensive transactional commerce with the environment.[2]

[2] von Bertalanffy's definition explicitly recognizes the first two of these criteria as present in all living organisms. A living organism, he says, is "an open system which continually gives up matter to the outer world and takes in matter from it, but which maintains itself in this continuous exchange in a steady state, or approaches such steady state in its variations in time" (1952a, p. 125). But elsewhere in this author's writing we find recognition of the additional criteria (1952a, p. 145; 1952b, p. 34).

While all of our theories view personality as an open system in some sense, still they can be fairly well classified according to the varying emphasis they place upon each of these criteria, and according to how many of the criteria they admit.

*Criterion 1*

Consider the first criterion of material and energy exchange. Stimulus-response theory in its purest form concentrates on this criterion to the virtual exclusion of all the others. It says in effect that a stimulus enters and a response is emitted. There is, of course, machinery for summation, storage, and delay, but the output is broadly commensurate with the intake. We need study only the two poles of stimulus and response with a minimum of concern for intervening processes. Methodological positivism goes one step further, saying in effect, that we do not need the concept of personality at all. We focus attention on our own measurable manipulations of input and on the measurable manipulations of output. Personality thus evaporates in a mist of method.

*Criterion 2*

The requirement of steady state for open systems is so widely admitted in personality theory that it needs little discussion. To satisfy needs, to reduce tension, to maintain equilibrium, comprise, in most theories, the basic formula of personality dynamics. Some authors, such as Stagner (1951) and Mowrer (1959) regard this formula as logically fitting in with Cannon's (1932) account of homeostasis.[3] Man's intricate adjustive behavior is simply an extension of the principle involved in temperature regulation, balance of blood

[3] In a recent review Mowrer (1959) strongly defends the homeostatic theory. He is distressed that the dean of American psychologists, Robert Woodworth (1958) has taken a firm stand against the "need-primacy" theory in favor of what he calls the "behavior-primacy" theory. With the detailed merits of the argument we are not here concerned. What concerns us at the moment is that the issue has been sharply joined. Need-primacy which Mowrer calls a "homeostatic" theory does not go beyond our first two criteria for an open system. Woodworth by insisting that contact with, and mastery of, the environment constitute a pervasive principle of motivation, recognizes the additional criteria.

volume, sugar content, and the like, in the face of environmental change. It is true that Toch and Hastorf (1955) warn against overextending the concept of homeostasis in personality theory. I myself doubt that Cannon would approve the extension, for to him the value of homeostasis lay in its capacity to free man for what he called "the priceless unessentials" of life (1932, p. 323). When biological equilibrium is attained the priceless unessentials take over and constitute the major part of human activity. Be that as it may, most current theories clearly regard personality as a *modus operandi* for restoring a steady state.

Psychoanalytic theories are of this order. According to Freud the ego strives to establish balance among the three "tyrants"—id, superego, and outer environment. Likewise the so-called mechanisms of ego defense are essentially maintainers of a steady state. Even a neurosis has the same basic adjustive function.[4]

To sum up: most current theories of personality take full account of two of the requirements of an open system. They allow interchange of matter and energy, and recognize the tendency of organisms to maintain an orderly arrangement of elements in a steady state. Thus they emphasize stability rather than growth, permanence rather than change, "uncertainty reduction" (information theory), and "coding" (cognitive theory) rather than creativity. In short, they emphasize being rather than becoming. Hence, most personality theories are biologistic in the sense that they ascribe to personality only the two features of an open system that are clearly present in all living organisms.

There are, however, two additional criteria, sometimes mentioned but seldom stressed by biologists themselves, and similarly neglected in much current personality theory.

[4] When we speak of the "function" of a neurosis we are reminded of the many theories of "functionalism" current in psychology and social science. Granted that the label, as Merton (1957) has shown, is a wide one, still we may safely say that the emphasis of functionalism is always on the usefulness of an activity in maintaining the "steady state" of a personality or social or cultural system. In short, "functional" theories stress maintenance of present direction allowing little room or none at all for departure and change.

*Transatlantic Perspective*

Before examining Criterion 3 which calls attention to the tendency of open systems to enhance their degree of order, let us glimpse our present theoretical situation in cross-cultural perspective. In this country our special field of study has come to be called "behavioral science" (a label now firmly stuck to us with the glue of the Ford millions). The very flavor of this term suggests that we are occupied with semiclosed systems. By his very name the "behavioral scientist" seems committed to study man more in terms of behavior than in terms of experience, more in terms of mathematical space and clock-time than in terms of existential space and time; in terms of response more than in terms of programing; in terms of tension reduction more than tension enhancement; in terms of reaction more than proaction.

Now let us leap our cultural stockade for a moment and listen to a bit of ancient Hindu wisdom. Most men, the Hindus say, have four central desires. To some extent, though only roughly, they correspond to the developmental stages of life. The first desire is for pleasure—a condition fully and extensively recognized in our Western theories of tension reduction, reinforcement, libido, and needs. The second desire is for success—likewise fully recognized and studied in our investigations of power, status, leadership, masculinity, and need-achievement. Now the third desire is to do one's duty and discharge one's responsibility. (It was Bismarck, not a Hindu, who said: "We are not in this world for pleasure but to do our damned duty.") Here our Western work begins to fade out. Excepting for some pale investigations of parental punishment in relation to the development of childhood conscience, we have little to offer on the "duty motive." Conscience we tend to regard as a reactive response to internalized punishment, thus confusing the past "must" of learning with the "ought" involved in programing our future (Allport, 1954, pp. 68–74). Finally, the Hindus tell us that in many people all these three motives pall, and they then seek intensely for a grade of understanding—for a philosophical or religious meaning—that will liberate them from pleasure, success, and duty (Smith, 1958). (Need I point out that most

Western personality theories treat the religious aspiration in reactive terms as an escape device, to be classified along with suicide, alcoholism, and neurosis?)

Now we retrace our steps from India to modern Vienna and encounter the existentialist school of logotherapy. Its founder, Viktor Frankl, emphasizes above all the central place of duty and meaning, the same two motives that the Hindus place highest in their hierarchy of desire. Frankl reached his position after a long and agonizing incarceration in Nazi concentration camps. With other prisoners he found himself stripped to naked existence (1959a). In such extremity what does a person need and want? Pleasure and success are out of the question. One wants to know the meaning of his suffering and to learn how as a responsible being he should acquit himself. Should he commit suicide? If so, why; if not, why not? The search for meaning becomes supreme.

Frankl is aware that his painfully achieved theory of motivation departs widely from most American theory, and he points out the implication of this fact for psychotherapy. He specifically criticizes the principle of homeostasis (1959b) as implying that personality is a quasiclosed system. To cater to the internal adjustments of a neurotic, or to assume that he will regain health by reshuffling his memories, defenses, or conditioned reflexes is ordinarily self-defeating. In many cases of neurosis only a total breakthrough to new horizons will turn the trick.

Neither Hindu psychology nor Frankl underestimates the role of pleasure and success in personality. Nor would Frankl abandon the hard won gains reflected in psychoanalytic and need theory. He says merely that in studying or treating a person we often find these essentially homeostatic formulations inadequate. A man normally wants to know the whys and wherefores. No other biological system does so; hence, man stands alone in that he possesses a degree of openness surpassing that of any other living system.

*Criterion 3*

Returning now to our main argument, we encounter a not inconsiderable array of theories that emphasize the tendency of human personality to go beyond steady states and to strive for an enhancement and elaboration of internal order even at the cost of considerable disequilibrium.

I cannot examine all of these nor name all the relevant authors. One could start with McDougall's proactive sentiment of self-regard which he viewed as organizing all behavior through a kind of "forward memory" (to use Gooddy's apt term—1959). Not too dissimilar is the stress that Combs and Snygg place on the enhancement of the phenomenal field. We may add Goldstein's conception of self-actualization as tending to enhance order in personality; also Maslow's theory of growth motives that supplement deficiency motives. One thinks of Jung's principle of individuation leading toward the achievement of a self (a goal never actually completed). Some theories, Bartlett and Cantril among them, put primary stress on the "pursuit of meaning." Certain developments in post-Freudian "ego psychology" belong here.[5] So too does existentialism with its recognition of the need for meaning and of the values of commitment. (The brain surgeon, Harvey Cushing, was speaking of open systems when he said: "The only way to endure life is to have a task to complete.")

No doubt we should add Woodworth's recent advocacy of the "behavior primacy" theory as opposed to the "need" theory, Robert White's emphasis on "competence," and Erikson's "search for identity."

Now these theories are by no means identical. The differences between them merit prolonged debate. I lump them here simply because all seem to me to recognize the third criterion of open systems, namely, the tendency of such systems to enhance their degree of order and become something more than at present they are.

We all know the objection to theories of this type. Methodologists with a taste for miniature and fractionated systems complain that they do not lead to "testable propositions" (cf. Roby, 1959). The challenge is valuable in so far as it calls for an expansion of research ingenuity. But the complaint is ill-advised if it demands that we return to quasiclosed systems

[5] Pumpian-Mindlin (1959) writes: "The focus of clinical psychoanalysis on ego psychology is a direct result of the change from a closed system to an open one" (p. 1051).

simply because they are more "researchable" and elegant. Our task is to study what is, and not what is immediately convenient.

*Criterion 4*

Now for our fourth and last criterion. Virtually all the theories I have mentioned up to now conceive of personality as something integumented, as residing within the skin. There are theorists (Kurt Lewin, Martin Buber, Gardner Murphy, and others) who challenge this view, considering it too closed. Murphy says that we overstress the separation of man from the context of his living. Experiments on sensory deprivation Hebb (1955) has interpreted as demonstrations of the constant dependence of inner stability on the flow of environmental stimulation. Why Western thought makes such a razor-sharp distinction between the person and all else is an interesting problem. Probably the personalistic emphasis in Judeo-Christian religion is an initial factor, and as Murphy (1958, p. 297) has pointed out the industrial and commercial revolutions further accentuated the role of individuality. Shinto philosophy, by contrast, regards the individual, society, and nature as forming the tripod of human existence. The individual as such does not stick out like a raw digit. He blends with nature and he blends with society. It is only the merger that can be profitably studied.

As Western theorists most of us, I dare say, hold the integumented view of the personality system. I myself do so. Others rebelling against the setting of self over against the world, have produced theories of personality written in terms of social interaction, role relations, situationism, or some variety of field theory. Still other writers, such as Talcott Parsons (1951) and F. H. Allport (1955), have admitted the validity of both the integumented personality system and systems of social interaction, and have spent much effort in harmonizing the two types of systems thus conceived.

This problem, without doubt, is the knottiest issue in contemporary social science. It is the issue which, up to now, has prevented us from agreeing on the proper way to reconcile psychological and sociocultural science.

In this matter my own position is on the conservative side. It is the duty of psychology, I think, to study the person-system, meaning thereby the attitudes, abilities, traits, trends, motives, and pathology of the individual—his cognitive styles, his sentiments, and individual moral nature and their interrelations. The justification is twofold: (*a*) there is a persistent though changing person-system in time, clearly delimited by birth and death; (*b*) we are immediately aware of the functioning of this system; our knowledge of it, though imperfect, is direct, whereas our knowledge of all other outside systems, including social systems, is deflected and often distorted by their necessary incorporation into our own apperceptions.

At the same time our work is incomplete unless we admit that each person possesses a *range* of abilities, attitudes, and motives that will be evoked by the different environments and situations he encounters. Hence, we need to understand cultural, class, and family constellations and traditions in order to know the schemata the person has probably interiorized in the course of his learning. But I hasten to warn that the study of cultural, class, family, or any other social system does not automatically illumine the person-system, for we have to know whether the individual has accepted, rejected, or remained uninfluenced by the social system in question. The fact that one plays the role of, say, teacher, salesman, or father is less important for the study of his personality than to know whether he likes or dislikes, and how he defines, the role. And yet at the same time unless we are students of sociocultural systems we shall never know what it is the person is accepting, rejecting, or redefining.

The provisional solution I would offer is the following: the personality theorist should be so well trained in social science that he can view the behavior of an individual as fitting any system of interaction; that is, he should be able to cast this behavior properly in the culture where it occurs, in its situational context, and in terms of role theory and field theory. At the same time he should not lose sight—as some theorists do—of the fact that there is an internal and subjective patterning of all these contextual acts. A traveler who moves from culture to culture, from situation to situation, is none the less a single person;

and within him one will find the nexus, the patterning, of the diverse experiences and memberships that constitute his personality.

Thus, I myself would not go so far as to advocate that personality be defined in terms of interaction, culture, or roles. Attempts to do so seem to me to smudge the concept of personality, and to represent a surrender of the psychologist's special assignment as a scientist. Let him be acquainted with all systems of interaction, but let him return always to the point where such systems converge and intersect and are patterned—in the single individual.

Hence, we accept the fourth (transactional) criterion of the open system, but with the firm warning that it must not be applied with so much enthusiasm that we lose the personality system altogether.

## General Systems Theory

There are those who see hope for the unification of science in what James Miller (1955) called "general behavior systems theory." This approach seeks formal identities between physical systems, the cell, the organ, the personality, small groups, the species, and society. Now critics (e.g., Buck, 1956) complain that all this is feeble analogizing, that formal identities probably do not exist, and that attempts to express analogies in terms of mathematical models result only in the vaguest generalities. As I see it, the danger in attempting to unify science in this manner lies in the inevitable approach from below, that is, in terms of physical and biological science. Closed systems or systems only partly open become our model, and if we are not careful, human personality in all its fullness is taken captive into some autistic paradise of methodology.

Besides neglecting the criteria of enhanced organization and transaction general systems theory has an added defect. The human person is, after all, the observer and interpreter of systems. This awkward fact has recently been haunting the founder of the operational movement, P. W. Bridgman (1959). Can we as scientists live subjectively within our system and at the same time take a valid objective view thereof?

Some years ago Elkin (1947) published the case of "Harry Holzer," and invited 39 specialists to offer their conceptualizations. As might be expected, many different conceptualizations resulted. No theorist was able entirely to divest the case of his own preconceptions. Each read the objective system in terms of the subjective. In other words, our theories of personality—all of them—reflect the temperament of the author fully as much as the personality of *alter*.

This sad spectre of observer contamination should not, I think, discourage us from the search for objectively valid theory. Truth, as the philosopher Charles Pierce has said, is the opinion which is fated to be ultimately agreed to by all who investigate. My point is that "the opinion fated to be ultimately agreed to by all who investigate" is not likely to be reached through a premature application of general systems theory, nor through devotion to any one partially closed theory. Theories of open-systems hold more promise, though at present they are not in agreement among themselves. But somewhere, sometime, I hope and believe, we shall establish a theory of the nature of personality which all wise men who investigate, including psychologists, will eventually accept.

## Some Examples

In the meantime, I suggest that we regard all sharp controversies in personality theory as probably arising from the two opposed points of view—the quasiclosed and the fully open.

The principle of reinforcement, to take one example, is commonly regarded as the cement that stamps in a response, as the glue that fixes personality at the level of past deeds. Now an open-system interpretation is very different. Feigl (1959, p. 117), for instance, has pointed out that reinforcement works primarily in a prospective sense. It is only from a *recognition* of consequences (not from the consequences themselves) that the human individual binds the past to the future and resolves to avoid punishment and to seek rewards in similar circumstances, provided, of course, that it is consonant with his interests and values to do so. Here we no longer assume that reinforcement stamps in, but that it is one factor among many to be considered in the programing of future action (Allport,

1946). In this example we see what a wide difference it makes whether we regard personality as a quasiclosed or open system.

The issue has its parallels in neurophysiology. How open is the nervous system? We know it is of a complexity so formidable that we have only an inkling as to how complex it may be. Yet one thing is certain, namely, that high level gating often controls and steers lower level processes. While we cannot tell exactly what we mean by "higher levels" they surely involve ideational schemata, intentions, and generic personality trends. They are instruments for programing, not merely for reacting. In the future we may confidently expect that the neurophysiology of programing and the psychology of proaction will draw together. Until they do so it is wise to hold lightly our self-closing metaphors of sowbug, switchboard, giant computor, and hydraulic pump.

Finally, an example from motivation theory. Some years ago I argued that motives may become functionally autonomous of their origins. (And one lives to regret one's brashness.)

Whatever its shortcomings the concept of functional autonomy succeeds in viewing personality as an open and changing system. As might be expected, criticism has come chiefly from those who prefer to view the personality system as quasiclosed. Some critics say that I am dealing only with occasional cases where the extinction of a habit system has failed to occur. This criticism, of course, begs the question, for the point at issue is why do some habit systems fail to extinguish when no longer reinforced? And why do some habit systems that were once instrumental get refashioned into interests and values having a motivational push?

The common counterargument holds that "secondary reinforcement" somehow miraculously sustains all the proactive goal-seeking of a mature person. The scientific ardor of Pasteur, the religio-political zeal of Gandhi, or for that matter, Aunt Sally's devotion to her needlework, are explained by hypothetical cross-conditioning that somehow substitutes for the primary reinforcement of primary drives. What is significant for our purposes is that these critics prefer the concept of secondary reinforcement, not because it is clearer, but because it holds our thinking within the frame of a quasiclosed (reactive) system.

Now is not the time to re-argue the matter, but I have been asked to hint at my present views. I would say first that the concept of functional autonomy has relevance even at the level of quasiclosed systems. There are now so many indications concerning feedback mechanisms, cortical self-stimulation, self-organizing systems, and the like (Chang, 1950; Hebb, 1949; Olds & Milner, 1954) that I believe we cannot deny the existence of self-sustaining circuit mechanisms which we can lump together under the rubric "perseverative functional autonomy."

But the major significance of the concept lies in a different direction, and presupposes the view that personality is an expanding system seeking progressively new levels of order and transaction. While drive motives remain fairly constant throughout life, existential motives do not. It is the very nature of an open system to achieve progressive levels of order through change in cognitive and motivational structure. Since in this case the causation is systemic we cannot hope to account for functional autonomy in terms of specific reinforcements. This condition I would call "propriate functional autonomy."

Both perseverative and propriate autonomy are, I think, indispensable conceptions. The one applies to the relatively closed part-systems within personality; the other to the continuously evolving structure of the whole.

A last example. It is characteristic of the quasiclosed system outlook that it is heavily nomothetic. It seeks response and homeostatic similarities among all personality systems (or, as in general behavior systems theory, among *all* systems). If, however, we elect the open system view we find ourselves forced in part toward the idiographic outlook. For now the vital question becomes "what makes the system hang together in any one person?" (cf. Taylor, 1958). Let me repeat this question, for it is the one that more than any other has haunted me over the years. *What makes the system cohere in any one person?* That this problem is pivotal, urgent, and relatively neglected, will be recognized by open-system theorists, even while it is downgraded and

evaded by those who prefer their systems semiclosed.

## FINAL WORD

If my discourse has seemed polemical I can only plead that personality theory lives by controversy. In this country we are fortunate that no single party line shackles our speculations. We are free to pursue any and all assumptions concerning the nature of man. The penalty we pay is that for the present we cannot expect personality *theory* to be cumulative—although, fortunately, to some extent personality *research* can be.

Theories, we know, are ideally derived from axioms, and if axioms are lacking, as in our field they are, from assumptions. But our assumptions regarding the nature of man range from the Adlerian to the Zilborgian, from the Lockean to the Leibnitzian, from the Freudian to the Hullian, and from the cybernetic to the existentialist. Some of us model man after the pigeon; others view his potentialities as many splendored. And there is no agreement in sight.

Nils Bohr's principle of complementarity contains a lesson for us. You recall that he showed that if we study the position of a particle we cannot at the same time study its momentum. Applied to our own work the principle tells us that if we focus on reaction we cannot simultaneously study proaction; if we measure one trait we cannot fix our attention on pattern; if we tackle a subsystem we lose the whole; if we pursue the whole we overlook the part-functioning. For the single investigator there seems to be no escape from this limitation. Our only hope is to overcome it by a complementarity of investigators and of theorists.

While I myself am partisan for the open system, I would shut no doors. (Some of my best friends are quasiclosed systematists.) If I argue for the open system I plead more strongly for the open mind. Our condemnation is reserved for that peculiar slavery to fashion which says that conventionality alone makes for scientific respectability. We still have much to learn from our creative fumblings with the open system. Among our students, I trust, there will be many adventurers. Shall we not teach them that in the pastures of science it is not only the sacred cows that can yield good scientific milk?

## REFERENCES


ALLPORT, F. H. *Theories of personality and the concept of structure.* New York: Wiley, 1955.

ALLPORT, G. W. Effect: A secondary principle of learning. *Psychol. Rev.*, 1946, **53**, 335–347.

ALLPORT, G. W. *Becoming: Basic considerations for a psychology of personality.* New Haven: Yale Univer. Press, 1954.

BRIDGMAN, P. W. *The way things are.* Cambridge: Harvard Univer. Press, 1959.

BRUNSWIK, E. The conceptual framework of psychology. In, *International Encyclopedia of Unified Science.* Vol. 1. Chicago: Chicago Univer. Press, 1955.

BUCK, R. C. On the logic of general behavior systems theory. In H. Feigl & M. Scriven (Eds.), *Minnesota studies in the philosophy of science.* Vol. 1. Minneapolis: Univer. Minnesota Press, 1956.

CANNON, W. B. *The wisdom of the body.* New York: Norton, 1932.

CHANG, H. T. The repetitive discharge of corticothalamic reverberating circuit. *J. Neurophysiol.*, 1950, **13**, 235–257.

ELKIN, F. Specialists interpret the case of Harry Holzer. *J. abnorm. soc. Psychol.*, 1947, **42**, 99–111.

FEIGL, H. Philosophical embarrassments of psychology. *Amer. Psychologist*, 1959, **14**, 115–128.

FRANKL, V. E. *From death camp to existentialism.* Boston: Beacon, 1959. (a)

FRANKL, V. E. Das homöostatische Prinzip und die dynamische Psychologie. *Z. Psychother. med. Psychol.*, 1959, **9**, 41–47. (b)

GOODDY, W. Two directions of memory. *J. indiv. Psychol.*, 1959, **15**, 83–88.

HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.

HEBB, D. O. The mammal and his environment. *Amer. J. Psychiat.*, 1955, **111**, 826–831. (Reprinted: In Eleanor Maccoby, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology.* New York: Holt, 1958. Pp. 335–341.)

MERTON, R. K. *Social theory and social structure.* (Rev. ed.) Glencoe, Ill.: Free Press, 1957.

MILLER, J. G. Toward a general theory for the behavioral sciences. *Amer. Psychologist*, 1955, **10**, 513–531.

MOWRER, O. H. A cognitive theory of dynamics. Review of R. S. Woodworth, *Dynamics of behavior. Contemp. Psychol.*, 1959, **4**, 129–133.

MURPHY, G. *Human potentialities.* New York: Basic Books, 1958.

OLDS, J., & MILNER, P. Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *J. comp. physiol. Psychol.*, 1954, **47**, 419–427.

PARSONS, T. *The social system.* Glencoe, Ill.: Free Press, 1951.

PUMPIAN-MINDLIN, E. Propositions concerning energetic-economic aspects of libido theory. *Ann. NY Acad. Sci.*, 1959, **76**, 1038–1052.

ROBY, T. B. An opinion on the construction of behavior theory. *Amer. Psychologist*, 1959, **14**, 127–134.

SMITH, H. *The religions of man*. New York: Harper, 1958. (Mentor Book Ed., 1959.)

STAGNER, R. Homeostasis as a unifying concept in personality theory. *Psychol. Rev.*, 1951, **58**, 5–17.

TAYLOR, J. G. Experimental design: A cloak for intellectual sterility. *Brit. J. Psychol.*, 1958, **49**, 106–116.

TOCH, H. H., & HASTORF, A. H. Homeostasis in psychology. *Psychiatry*, 1955, **18**, 81–91.

VON BERTALANFFY, L. *Problems of life*. (Trans. of *Das biologische Weltbild*, 1949) New York: Wiley, 1952. (a)

VON BERTALANFFY, L. Theoretical models in biology and psychology. In D. Krech & G. S. Klein (Eds.), *Theoretical models and personality theory*. Durham: Duke Univer. Press, 1952. (b)

WOODWORTH, R. S. *Dynamics of behavior*. New York: Holt, 1958.

(Received September 9, 1959)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 311-318

# ON THE RELIABILITY OF THE ANAMNESTIC INTERVIEW[1]

ERNEST A. HAGGARD[2]
*College of Medicine, University of Illinois*

ARNE BREKSTAD AND ÅSE GRUDA SKARD
*Psychological Institute, University of Oslo*

THE anamnestic interview is used extensively by psychiatrists, guidance counselors, anthropologists, and other persons who are interested in the prior events in the life of another person. This procedure, which relies upon the memory of an informant, is used so often simply because the easiest way (and often the only way) to obtain many types of information about another person's background is to ask him. And, because the preoccupation of the interviewer is to obtain "the truth," he generally is inclined to accept with gratitude whatever the informant tells him, and to reify it.

Many investigators have questioned the extent to which anamnestic material gives a true picture of prior events. Although the vast literature on memory and its fallibility bears on this question, relatively few studies have been carried out that deal with it directly. The purpose of this study is to inquire about the reliability of various types of anamnestic information and to throw light on some of the factors that may distort an informant's report during an anamnestic interview.

## METHOD

*Data and sample.* The data are based on the records of a longitudinal study of the families of 19 children born and reared in Oslo and on a follow-up anamnestic interview with the mothers of these children. Over the duration of this study numerous interview and observational data have been collected on the parents and, in addition, many types of measurement, test, and observational data have been collected on the children. These data were collected for several purposes, one of which was to provide the opportunity for later investigating the reliability of the anamnestic interview.[3]

The sample was drawn from 55 working class families who visited a well-baby clinic in Oslo in 1952–53. Of the 19 families, 12 were to have their first child (4 were boys, 8 were girls) and 7 were to have their second child (6 were boys, 1 was a girl). Since the parents were interviewed before the children were born and at successive intervals thereafter, it was possible to sample several time periods and several areas of information or types of scales. The children were between 7 and 8 years of age at the time of the anamnestic interview.

*Sampling time periods.* Three time samples were selected for study: the prenatal period (about a month before delivery), during the child's first year, and during the child's sixth year. Thus, whereas the prenatal scales have to do primarily with the parents' projected attitudes and expectations, the first and sixth year scales are tied more closely to their experiences with their child and what they expected from him at that age.

From 18 scales available from the larger project for the prenatal period, 10 were selected; from 31 scales available for the first year period, 6 were selected; and from 36 scales available for the sixth year period, 11 were selected. The 27 scales used in this study are given in Table 1 below.

With respect to the possible effect of time on the reliability of memory, two samples of scales covered an interval of from 1 to 2 years and two samples of scales covered an interval of from 7 to 8 years. For the former scales, one sample (Scales 1–3) covered an interval of about 1 year (these three scales were readministered during the first year and were anamnestic to the prenatal period); the other sample (Scales 17–27) covered an interval of from 1 to 2 years. For the longer intervals, one sample (Scales 1–10) covered an interval of about 8 years and the other sample (Scales 11–16) covered an interval of about 7 years (see Table 1).

*Sampling types of information (scales).* Five criteria were used to select the scales for this study: (*a*) that, where possible, they should be representative of the questions which guidance counselors, clinical psychologists, etc. typically ask mothers in an anamnestic interview; (*b*) that it would not be unreasonable to ask the mothers for such information in view of their

[1] This study was supported by Research Grant M-1737 from the National Institute of Mental Health, of the National Institutes of Health, United States Public Health Service.

[2] On leave from the Department of Psychiatry, University of Illinois during 1959–60 as a Public Health Service Research Fellow of the National Institute of Mental Health.

[3] This project, entitled "Children's Personality Development in an Oslo Milieu," was begun in 1950 under the direction of Åse G. Skard. A full description of this project is available in: Skard, Åse G. Scope and methods of the Oslo project. 1960. (Mimeo)

previous contacts with the project; (*c*) that the scales should cover a variety of types of information; (*d*) that the project records should contain sufficiently consistent and precise information on the 19 cases to minimize uncertainty in our scaling procedure;[4] and (*e*) where possible, the scales should not be highly intercorrelated.[5]

*Scaling procedure.* After examining the protocols, we decided to use a normalized six-point scale to quantify the data. For convenience, we will call the scales based on the project protocols "historical" and the corresponding scales based on the follow-up interview "anamnestic."

Although the historical scales for the prenatal and first year periods were originally in the form of ratings, it was necessary to recode them to have scales of uniform size and explicit definition. The historical scales for the sixth year period were originally in the form of ranks based on the independent judgments previously made by two other raters.[6] Another difference between the earlier and later scales lies in the fact that the prenatal and first year scales were generally based on specific questions to the mothers, whereas the sixth year scales were more often inferred from the "general tone" of the historical material. For the sixth year period, then, it was not possible to have completely parallel scoring procedures, since in the anamnestic interview, the questions were phrased to bear directly on Scales 17 to 27.

For the historical scales, the extremes and midpoints were anchored by excerpts from the protocols, the cases coded along a semiquantitative graphic continuum (see Champney, 1941), and later converted into a normalized six-point scale. The same person who established the mid- and end-points on the historical scales also coded the anamnestic scales. Each historical scale was coded also in terms of whether the mother had shown marked anxiety in connection with it at the time. From four to eight mothers were sufficiently anxious in connection with the content of Scales 1, 4, 5, and 7 to permit later analysis of the possible effects of such disturbances on the reliability of recall.

In coding the anamnestic scales we wished to obtain some measure of the extent of any random discrepancies or systematic shifts in the mothers' responses to the questions corresponding to each of the scales. Thus, on each anamnestic scale, each mother's response was coded in terms of where it would have fallen on the corresponding historical scale. Because discrepancies did occur between the mothers' reports in the historical and anamnestic interviews, 4 of the 513 scores on the anamnestic scales fell beyond the 1–6 point range. In these instances we decided to use the scale value of 1 or 6 to preserve the uniform size of the scales. Also, because of changes or shifts in the mothers' reports, not all of the anamnestic scales were distributed normally, although the distribution over all of the scales was approximately normal.

*The anamnestic interview.* The interview itself was justified to the mothers in terms of our interest in "how they see things now" and our desire to learn how the children were doing in school (since, in Norway, children begin school in their seventh year). Although the mothers had been interviewed many times by members of the project staff, they were not acquainted with the present interviewer.

An interview guide was drawn up around the 27 scales with the addition of questions to fill in time gaps and to cover various points of interest to the project staff. Although this interview, which usually lasted a little over 1 hour, was designed to be a representative anamnestic interview, it was atypical in three respects: it was recorded; the mothers knew that the project files contained their previous responses to similar questions; and the mothers did not come to the interview because they were disturbed about, and hence seeking help for, some "problem." If these three differences have introduced any bias into our findings, their effect has probably been to elicit more accurate or reliable responses than would have been the case if the mothers had been emotionally disturbed or had had reason to give an unduly favorable (or unfavorable) account to the interviewer.

*Methods of data analysis.* The coefficient of intraclass correlation, *R*, was used as the most sensitive and accurate measure of reliability (see Haggard, 1958) to reflect the relative congruence of the mothers' responses to the questions underlying the historical and anamnestic scales. Appropriate coefficients were computed over scales and over individuals, and frequently these *R*s were averaged (via the *z* transformation) to provide measures for groups of scales or individuals.

In analyzing these data, our primary interest has been to examine the extent to which various factors or conditions have contributed to the reliable (or distorted) recall in an anamnestic interview and to indicate their relative importance rather than to stress the statistical significance of our findings or to infer that such factors will have comparable influence in all similar contexts.

## RESULTS

### *Validity or Reliability?*

In order to speak of validity in the customary sense, one would need objective external criteria for the scales. Such criteria were available for Scales 11–14 since each child's length, weight, and time of birth were available from hospital records and the approximate length of breast-feeding was available from clinic records and frequent visits to the home during the first year. The *R*s for these four scales ranged

[4] For one case, which had dropped out of the study, 3's and 4's were assigned at random for the 11 sixth year scales (17–27). Otherwise, the protocols contained adequate information for assigning the remaining 502 (of the 513) historical scores.

[5] For the sixth year period, for example, the rank order intercorrelations among the 11 scales ranged from −.39 to .55, with a median $\rho$ of .03.

[6] The interjudge $\rho$'s ranged from .50 to .88 before the two rankings were combined to form the basis for Scales 17–27.

TABLE 1

RELIABILITY COEFFICIENTS BETWEEN HISTORICAL AND ANAMNESTIC REPORTS

| Prenatal scales: | Six-month anamnesis | Eight-year anamnesis |
|---|---|---|
| 1. Attitude toward being pregnant | .02 | .60 |
| 2. Strength of wish for a certain sex | .50 | .23 |
| 3. What sex did the mother wish for | .81 | .59 |
| 4. Did the parents plan to have the child | | .83 |
| 5. Pregnancy difficulties | | .40 |
| 6. Anxiety for the childbirth | | .09 |
| 7. Anxiety for the child's normality and appearance | | .32 |
| 8. Attitude toward regulating child's food schedule | | .41 |
| 9. Attitude toward obedience | | .37 |
| 10. Attitude toward punishment as discipline | | .73 |
| | $R_{av.}$ = | .49 |
| **First year scales:** | ***R* from measures** | |
| 11. Length of child at birth | .91 | .89 |
| 12. Weight of child at birth | .77 | .71 |
| 13. Child born at normal time | .91 | .81 |
| 14. Length of breast-feeding | .96 | .77 |
| 15. Attitude toward the delivery (childbirth) | | .56 |
| 16. Attitude toward regulating child's food schedule | | .46 |
| | $R_{av.}$ = | .73 |
| **Sixth year scales:** | | |
| 17. Security (or lack of anxiety) in child care | | .49 |
| 18. Agreement between the parents in child care | | .22 |
| 19. Encouraging the child to dress alone | | .13 |
| 20. Encouraging the child to wash himself alone | | .35 |
| 21. Encouraging the child to eat alone | | .25 |
| 22. Encouraging the child to go to the toilet alone | | .17 |
| 23. Giving child freedom to leave home alone | | .53 |
| 24. Letting child decide things by himself | | −.09 |
| 25. Attitude toward aggression in general | | .35 |
| 26. Attitude toward child's aggression to other children | | .18 |
| 27. Letting child choose his own playmates | | −.08 |
| | $R_{av.}$ = | .24 |
| | $R_{av.}$ (total interview) = | .47 |

from .71 to .89, with an average $R$ ($R_{av.}$) of .80.[7] It is worth noting that this group of four scales contains the highest *R*s in Table 1.

Even though the interviewer may be seeking "the truth" in conducting an anamnestic interview, in most cases it is meaningless from a research point of view to phrase the matter in these terms, simply because there is no way of telling what was "true" at the time the event in question took place. One could, of course, assume that the informant's report at the time should be taken at face value but, even so, this assumption would hardly enable one to speak of "validity." With respect to Scale 6, for example, the mother stated on a given occasion how anxious she was about the coming childbirth—but who knows whether she was also trying to be brave or to seek sympathy, or whether her statements on another day (and in a different frame of mind) would have been the same?

Although we may speak of the findings for Scales 11–14 as validity coefficients, on the remaining 23 scales we will limit our interest to the degree of congruence between the historical and anamnestic reports, as indicated by *R*s, in order to examine various types of factors associated with the degree of agreement between these two sets of reports. More specifically, we will focus our attention on four types of conditions that appear to influence the reliability of the mothers' reports during the anamnestic interview. These conditions involve the length of the time interval between the historical and anamnestic interviews, differences in the type of information underlying the scales, sex and birth-order differences among the children, and individual differences among the mothers.

### *Time Intervals*

Psychologists ordinarily assume that more distant events will be recalled less well than more recent events—with some exceptions, of course. In this study it was possible to group the 27 scales into those which involved a relatively short time interval (1 to 2 years) and those which involved a relatively long interval (7 to 8 years). There were two samples of scales in each grouping as follows:

1. Short interval: prenatal to first year. Scales 1–3 were based on questions asked during the first year and were anamnestic to the prenatal Questions 1–3. The *R*s for these three scales were .02, .50, and .81, with $R_{av.}$ = .51. The fact that the *R* of .02 for Scale 1 (attitude toward being pregnant) is so much lower than the others suggests that a factor such as anxiety might have influenced the mothers' reports to this question during the first year. (This suggestion is strengthened by the fact that, when asked the same question in the recent anam-

[7] The corresponding *R*s based on the physical measures were .91, .77, .91, and .96 for Scales 11–14, respectively, with $R_{av.}$ = .91. The difference between these two sets of *R*s is due to the greater precision of the original measures.

nestic interview, the *R* between these two reports, which covered an interval of about 8 years, was .60.)

2. Short interval: sixth year to anamnestic interview. Scales 17–27 are based on questions regarding parental reports made when the children were 6 years old. The *R*s for these scales range from −.09 to .53, with $R_{av.} = .24$. From Table 1 it is clear that this group of scales contains the lowest *R*s, which finding we would not expect on the basis of the time factor alone.

There are at least two possible types of explanation for the generally low and variable *R*s in this group. First, as we have already indicated, the interviewing and scoring procedures for the sixth year scales were not always strictly comparable for the historical and anamnestic scales. A second possibility is that the sixth year scales involved attitudes and criteria for the child's behavior which were themselves in a state of flux and perhaps were not distant enough in time to become stabilized. Thus, it may have been difficult for the mothers to differentiate their norms for their children at age 6 and at the time of the follow-up interview when they were from 7 to 8 years old. From this point of view, one might reason that the mothers were caught up in gradual changes in (or current uncertainties or conflicts about) their norms with respect to, for example, the extent to which the child should determine his own behavior (Scale 24) or choose his own playmates (Scale 27).

3. Long interval: prenatal to anamnestic interview. The *R*s for Scales 1–10 range from .09 to .83, with $R_{av.} = .49$. This $R_{av.}$ would have been somewhat higher (.55) if the two anxiety scales (6 and 7) were excluded. Even so, the $R_{av.}$ of .49 for this 8-year interval is appreciably higher than the above $R_{av.}$ of .24 for the more recent, sixth year scales.

4. Long interval: first year to anamnestic interview. The *R*s for Scales 11–16 range from .46 to .89, with $R_{av.} = .73$. This group includes, of course, the four "validity" scales; otherwise the *R*s of .56 and .46 for Scales 15 and 16 approximate the $R_{av.}$ of .49 for the above group of long-interval scales.

One may speculate as to why, generally speaking, the coefficients indicating the relative consistency or agreement between the mothers' reports on two different occasions tend to be higher for the longer time intervals. But speculations aside, an inspection of Table 1 suggests that the type of scale involved played a more important role in determining the size of the coefficients than did the length of the time interval.

### *Types of Scales*

Because of uncertainties regarding the comparability of the historical and anamnestic scales for the sixth year period, we will restrict our discussion of the various types of scales to those for which such questions do not arise, namely, Scales 1–16. These scales can be grouped under five headings, as follows:

1. "Hard facts." In our earlier discussion of the question of validity, we saw that the $R_{av.}$ for Scales 11–14 was .80, and that the coefficients for this group were higher than for any other grouping in spite of the rather long time interval. Perhaps the mothers were able to recall these facts so accurately because they were quantitative and relatively independent of their own attitudes, wishes, or anxieties.

2. General wishes and attitudes. Eight scales (2, 3, 4, 8, 9, 10, 15, and 16) had to do with the parents' wishes and attitudes, as to whether the child was wanted and planned for, the desired sex of the child, and, generally speaking, their attitudes toward various aspects of child rearing. The *R*s for these eight scales ranged from .23 to .73, with $R_{av.} = .50$.

3. Anxiety scales. We may assume that during the prenatal period all the mothers were more or less anxious about their own and their child's well-being during the childbirth and whether their child would be healthy and undeformed. The *R*s for Scales 6 and 7 were .09 and .32, with $R_{av} = .21$. Compared with the above two groupings, the *R*s for these scales suggest that anxiety interfered with the mothers' accurate recall of their earlier attitudes regarding these matters.

4. Anxiety ratings on attitude scales. In rating the early historical material the statements of each mother relevant to each scale were also coded in terms of whether they touched upon areas or experiences that were markedly disturbing to her. On three scales (1, 4, and 7) there were 8, 6, and 4 of the 19 mothers, respectively, who were judged to show clear signs of anxiety. The *R*s for these subgroups were .03 on Scale 1, .08 on Scale 4, and

−.37 on Scale 7, with an $R_{av.}$ of −.09 for the three scales. The corresponding *R*s for the mothers who were judged not to have been very anxious in connection with these matters were .17 on Scale 1, .56 for Scale 4, and .37 for Scale 7, with an $R_{av.}$ of .38.[8] The two sets of *R*s are in line with our previous findings that anxiety interferes with the accurate recall of this type of material.

5. Anxiety ratings on a nonattitude scale. The only other scale from the early material on which a sufficient number of mothers were judged to have shown clear signs of anxiety was Scale 5. The five mothers who expressed anxiety at the time about their difficulties during pregnancy did tend to recall their difficulties during the anamnestic interview ($R = .32$), whereas the nonanxious mothers tended not to do so ($R = -.37$). The apparent differential effect of anxiety on the reliability of recall of attitudes and of experiences will be touched upon again when we consider whether the mothers were having their first child or had had one previously.

*Birth Order and Sex Differences among Children*

As we observed earlier, the sample of families was selected for study before the birth of the children who also were to be studied over a period of several years. Under such conditions if, for practical reasons, one may not be able to reject the families initially selected, one runs the risk of getting a sample of children that is not balanced between boys and girls. Unfortunately, in the present study the birth order and sex of the children are confounded: first-born tends to be synonymous with girls and second-born tends to be synonymous with boys. Specifically, 8 of the 12 first-born children are girls and 5 of the 7 second-born children are boys.

*Birth order of child.* In discussing our findings with respect to birth order differences, we will limit our attention to two prenatal scales (5 and 6), the responses to which should not have been influenced by the sex of the unborn child, and to the 11 sixth year scales involving recent socialization attitudes.

[8] The *R*s for all 19 mothers taken together were .60, .83, and .32 for Scales 1, 4, and 7, respectively, with an $R_{av.}$ of .63. The lower *R*s for both the anxious and nonanxious mothers is due to the fact that these two subgroups tended to be distributed at different ends of the six-point scales, so that *R* is raised when the subgroups are combined.

On Scale 5, the 12 mothers who had their first child recalled their difficulties during pregnancy more accurately ($R = .57$) than the 7 mothers who had their second child ($R = .02$). On the other hand, on Scale 6 the mothers who had their first child tended not to recall their earlier reported anxiety over the childbirth process ($R = -.31$), as opposed to the mothers who had their second child ($R = .42$). These findings recall our previous discussion of Scales 5 and 6, where we cited data suggesting that anxiety served to facilitate accurate recall when associated with the mothers' experiences, but interfered with such recall when associated with earlier reported attitudes. One cannot be certain that mothers about to have their first child were more "anxious" than the others; they may have been only more involved and attentive to their pregnancy than the mothers who had been through the experience successfully once before. But whether it was anxiety or attentiveness, it appears that such factors influence differentially the recall of prior experiences and attitudes.

There were also differences in the reliability coefficients of mothers having their first or second child on the sixth year scales. The 12 mothers in the first group had an $R_{av.}$ of .10 for these 11 scales, whereas the corresponding $R_{av.}$ for the second group was .29. Furthermore, the individual *R*s were lower on 9 of the 11 scales for the mothers who had their first child. These findings are in line with the suggestion that the mothers' possible confusion or uncertainty about currently changing norms might interfere with the accurate recall of their previous attitudes in connection with this set of scales.

*Sex of child.* One may inquire as to whether mothers tended to give more reliable reports during the anamnestic interview, depending on whether their child was a boy or a girl—that is, whether it was of the opposite or the same sex as her own. Reasons for such an inquiry lie in the possible existence of culturally held attitudes that favor children of one sex more than the other and/or of less consciously held attitudes that favor children of the same or the opposite sex. But, because birth order and sex are confounded in this sample, we must consider any sex differences within the first- and second-born groups.

The relevant findings are summarized in

TABLE 2
AVERAGE RELIABILITIES FOR MOTHERS OF FIRST- AND SECOND-BORN BOYS AND GIRLS

| Scales | First-born | | Second-born | |
|---|---|---|---|---|
| | Boy ($n = 4$) | Girl ($n = 8$) | Boy ($n = 6$) | Girl ($n = 1$) |
| Prenatal (Scales 1–10) | .36 | .53 | .47 | (.68) |
| First year (Scales 11–16) | .57 | .71 | .80 | (.16) |
| Sixth year (Scales 17–27) | .02 | .18 | .03 | (.45) |
| Total interview (Scales 1–27) | .32 | .46 | .44 | (.54) |

Table 2. Here we find that, for the first child, the mothers of girls tended to give more reliable reports during the anamnestic interview than did the mothers of boys. On the second child, however, the coefficients for the mothers of boys are about the same as those for mothers of first-born girls. But because of the small number of cases in each of these subgroups and other complicating factors,[9] these findings must be thought of as suggesting that birth order and sex differences might influence the recall of prior events in an anamnestic interview.

*Individual Differences among Mothers*

Finally, we can inquire about possible differences among the mothers in the extent to which their reports during the anamnestic interview agreed with what they had said earlier. In this connection, reliability coefficients indicating the agreement in the mothers' reports on the 27 historical and anamnestic scales were computed. These *R*s range from .04 to .67, with $R_{av.} = .47$. Thus, although some mothers gave anamnestic reports that agreed reasonably well with what they had said previously, none of the mothers showed the degree of accuracy, or reliability, that one would hope for. Nor can it be overlooked that the anamnestic reports of some mothers were essentially random, at least within the limits of our scales. Furthermore, we find even greater variation among the mothers when we compute such reliability coefficients for the scales in each of the three time periods. Here we find that the *R*s ranged from −.39 to .91 for the 10 prenatal scales, from −.14 to 1.00 for the 6 first year scales, and from −.42 to .55 for the 11 sixth year scales. Even allowing for the fact that these latter *R*s may be relatively unstable because they are based on a small number of scales, one is nonetheless struck with the general *in*accuracy of the anamnestic reports given by many of the mothers.

The frequent low (or negative) coefficients and the wide variation among them raises the following question: Did the mothers tend to be consistently accurate or inaccurate in their reports involving different groups of scales, such as those in the three time periods? An answer to this question was arrived at by converting the 57 *R*s of the 19 mothers for each of the three time periods to *z* scores (to obtain normalized scores) and then computing an *R* for the 19 mothers over the three time periods taken together. This *R* was −.38, which clearly indicates the mothers' inconsistency in the reliability of their reports over these periods. In other words, if the mothers tended to recall the scales for one time period more accurately than the average, they tended to recall the scales for the other time periods less accurately. This tendency is also seen when we correlate the above *z* scores for each pair of time periods. Here we find that the *R* is −.07 between the prenatal and first year periods, is −.54 between the first year and sixth year periods, and is −.23 between the prenatal and sixth year periods.[10]

Since the implications of these findings are all too obvious, we were led to inquire whether

[9] These subgroups also differ in several other ways, the relevance of which is not always clear for this study. For example, the parents of second-born children were more secure in their treatment of the child, the parents of girls were less well-adjusted to each other, even before the child was born; the mothers breast-fed their boys about twice as long as the girls and also tended to treat them better in various other ways as well. These and other birth order and sex differences in this sample are discussed in detail in a paper by Eckhoff, Gauslaa, and Baldwin (in press).

[10] The disconcerting finding that informants in an anamnestic interview tend not to give equally accurate reports for all scales has also been reported by other investigators using larger samples of subjects. In one study (Pyles, Stoltz, & Macfarlane, 1935), which was based on the reports of 252 mothers, the intercorrelations (*r*'s) ranged from −.18 to .27 among five scales: month walked alone, month of first tooth, birthweight, weight at 12 months, and duration of labor. (We would have called the first four of the above "hard fact" scales; the authors report the highest *r*'s—of .22 and .27—among the first three of the above scales. Also, the scales used in the present study cover a broader range than those cited above, which fact may account for the greater variation in the reliability coefficients.)

an interviewer could tell ahead of time which informants would be the "reliable" reporters of which class of prior events in an anamnestic interview. We first asked two persons who had interviewed the mothers and also were familiar with the project records to see if they could formulate on the basis of their subjective impressions any clues which might be helpful in this respect. Using this approach, they were not able to differentiate among the mothers in terms of the reliability coefficients for any of the time periods or for all 27 scales taken together.

As a second approach, and on the basis of our other findings, we inquired whether the factor of anxiety in either its general or specific aspects might be related to the differences among the mothers in the accuracy of their recall. Global anxiety ratings by the sixth year and anamnestic interviewers were not significantly related to the above *R*s, nor were the scores on the anxiety scales. In the latter case, we combined Scales 6 and 7 from the prenatal period (anxiety over the child's appearance and childbirth) and Scale 17 from the sixth year period (security—or its converse, anxiety—in rearing the child). These scales also did not differentiate significantly among the mothers with high and low $\bar{R}$s for any of the time periods or for the total set of anamnestic scales. However, the fact that we were not able by these procedures to find indices which are related to the mothers' reliability coefficients does not mean that such indices do not exist.

## DISCUSSION

The findings of this study indicate that, in many respects, the mothers' statements during the anamnestic interview were not particularly accurate as reports of prior events. It appears that the anamnestic material did not reflect the mothers' earlier experiences and attitudes so much as their current picture of the past. This seems obvious enough. What did not seem obvious when we began this study is that the "reliability" of their picture of the past varied so greatly for different types of anamnestic material and from one mother to another. These findings do not, however, alter the fact that anamnestic interviews are usually conducted by persons who want to learn about what actually happened in the past.

An additional complication is the evidence that anxiety serves to distort the accurate recall of many experiences. In coming to a school or other guidance clinic, for example, the informant usually is anxious and unable to cope with some kind of problem, and so asks for help. In situations of this sort the interviewer inquires about the past so that he can make intelligent decisions about what kind of help should be given. It is an ironic possibility that in those situations where accurate information is most needed, it may be most distorted.

In this connection it is worth noting that the anamnestic interview, or some variation of it, has been used extensively to build up a body of "data" regarding, for example, the genetic roots of adult character structure. Since such "data" are frequently obtained from adult patients in psychotherapy, one may wonder about the extent to which they give an accurate, representative picture of the genesis of adult character structure in general. Perhaps the fact that many current theories of personality have been based on the reports of disturbed persons in therapy accounts in part at least for their predominant emphasis on conflicts and maladaptive mechanisms.

In any case, since the anamnestic interview is such an important practical and theoretical tool, its strengths and limitations should be investigated more thoroughly than we have been able to do. Even though the present study is ancillary to a broader project which was designed to meet other goals as well, we have been able to investigate the effects of certain conditions or sources of variation in anamnestic reports. On the basis of our findings it seems clear that more adequate investigations of the anamnestic interview can and should be carried out. To be most fruitful such studies should be designed with certain sampling, interview technique, and data analysis considerations in mind ahead of time.

*Sampling.* Because of the wide variety of situations in which the anamnestic interview is used, it would be desirable to use an experimental design with representative sampling of different types of informants, contents (attitudes, experiences, facts), etc. in different settings. Where possible, additional information should be collected to investigate, first, the "validity" as well as the "reliability" of anamnestic reports and, second, to identify factors that may be related to the marked individual differences among informants in the extent to

which they recount accurately certain types of information.

*Interview technique.* In this study we have only alluded to problems of interview technique. Particular care should be taken in eliciting anamnestic material since what the informant says depends in part on the relationship between him and the interviewer and by how the questions are phrased by the one and interpreted by the other. In this connection, it is possible that some of our reliability coefficients would have been substantially higher if the project as a whole had not also attempted to achieve a variety of other goals. For example, while collecting the historical material, the mothers were sometimes permitted to digress from the questions to discuss related matters of concern to them. Although this procedure served other purposes of the project, it complicated this particular study, and was one of the reasons why the historical and anamnestic materials from the sixth year period are not strictly comparable and hence could not be coded with satisfactory precision.

*Data analysis.* Although indices such as reliability coefficients give a general idea of how this or that variable contributes to or interferes with the accuracy of anamnestic reports, this approach is relatively limited in the amount of information it provides. Given the proper experimental design, sampling procedures, and scaling techniques, one could approach the problem in terms of the major sources of variation (e.g., components of variance) and their interactions in anamnestic reports. This type of information would contribute more effectively to a better understanding of memory processes and also would enable the interviewer to estimate the relative accuracy of various types of anamnestic material obtained under specified conditions.

## SUMMARY

This research was designed to study the effects of some of the factors which may influence the reliability of anamnestic reports. From a longitudinal study, records of 19 mothers' interviews before the children were born, when they were 1 and 6 years old, and other records were used to provide 27 normalized six-point scales. The scales were selected to cover relatively short (1 to 2 year) and long (7 to 8 year) time intervals and several different types of information. When the children were between 7 and 8 years old, the mothers were given an anamnestic interview which was built around the same 27 scales. The coefficient of intraclass correlation was used to give an index of the agreement or reliability of the mothers' reports between the earlier and the anamnestic interviews.

In general terms, we found that the mothers' statements during the anamnestic interview were not particularly accurate as reports of prior events. It appears that the anamnestic material did not reflect their earlier experiences and attitudes so much as their current picture of the past.

In more specific terms, we found that:

The length of time interval, by itself, was not meaningfully related to the reliability of the anamnestic reports.

Clear differences occurred from one type of information to another. The "hard fact" data, such as the length of child at birth were recalled best of all; general wishes and attitudes (not involving anxiety) were recalled next best; and attitude scales, concerning which of the mothers were anxious at the time, were recalled least well, with reliabilities around zero. If anxiety was associated with actual experiences, however, it seemed to facilitate the accurate recall of those experiences.

There were large individual differences among the mothers in the reliability of their anamnestic reports; furthermore, they were not consistent in the accuracy of their reports from one group of scales to another.

## REFERENCES

CHAMPNEY, H. The measurement of parent behavior. *Child Develpm.*, 1941, **12**, 131–166.

ECKHOFF, EVA, GAUSLAA, J., & BALDWIN, A. L. Parental behavior toward boys and girls of preschool age. *Acta psychol., Amsterdam*, in press.

HAGGARD, E. A. *Intraclass correlation and the analysis of variance.* New York: Holt-Dryden, 1958.

PYLES, M. K., STOLTZ, H. R., & MACFARLANE, J. W. The accuracy of mothers' reports on birth and development data. *Child Develpm.*, 1935, **6**, 165–176.

(Early publication received June 15, 1960)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 319-322

# ACQUIESCENT RESPONSE SET AND THE PSYCHASTHENIA SCALE: AN ANALYSIS VIA THE *AUSSAGE* EXPERIMENT

JAMES L. KUETHE
*Johns Hopkins University*

"I object. Counsel is attempting to lead the witness." This cry is familiar to everyone as part of the drama of the courtroom. Judges and lawyers have long been aware that the testimony of a witness may be biased if "leading questions" are permitted. Witnesses do not usually intend to give false information to a leading question but rather seem to be unaware that their testimony has been influenced by the examiner.

There have been two main approaches to the study of errors of testimony, the investigation of errors made during free recall and the investigation of errors induced by an examiner. A good review of early research on the topic is that of Whipple (1909). The *Aussage* demonstration has often been used in the classroom as a dramatic means of illustrating the low reliability of testimony. The typical procedure is for the instructor to have a confederate come up in front of the class and behave in some unusual manner. The class is then questioned about the events that have just taken place. The variability of the reports received convinces the class that testimony is subject to considerable error. This interesting pedagogical device only touches on one aspect of the problem of error of testimony.

The form of a question plays a major role in introducing error tendencies in the response (Burtt & Gaskill, 1932; Muscio, 1916), hence the legal concern aroused by "leading questions." However, there are considerable individual differences in susceptibility to influence of this nature that are ignored in studies interested only in showing that free recall is not reliable. The testimony of some persons is far more reliable than that of others. It was the aim of the present investigation to examine personality variables associated with individual differences in the accuracy of testimony.

Acquiescence to suggestion would seem to be involved when report is influenced by the form of a question. It is also obvious that most persons would not be as dramatically influenced if they were aware that pressures were being exerted upon them. Thus two personality variables appear to be important determinants of the extent to which a person's testimony may be influenced: an acquiescent response set and a tendency to be unaware of influence.

Kuethe (1959) has shown that persons with a strong acquiescent response set tend to cooperate with the experimenter and follow his instructions to perform a long uninteresting task. The acquiescent response set in this way relates to a tendency to comply with the dictates of another individual.

Eriksen, Kuethe, and Sullivan (1958) found that awareness of influence in verbal learning is related to scores on the Psychasthenia (*Pt*) scale of the MMPI. Low scorers on the *Pt* scale tended to lack verbal awareness of influence in a manner consistent with clinical theory and observation.

It was the hypothesis of the present investigation that persons both strong in acquiescent response set and low in score on the *Pt* scale would be more subject to influence when giving testimony than individuals not showing this personality pattern.

## METHOD

The subjects were 230 students enrolled in an introductory course at Teachers College, Columbia University. During a class meeting all subjects were given the *Pt* scale of the MMPI and a measure of the acquiescent response set. The measure of the acquiescent or positive response set was essentially a measure of the tendency to agree to suggested concepts and is described in detail by Kuethe (1959).

Several weeks later a classroom "drama" was staged to reveal individual differences in fidelity of testimony. The sequence of events was as follows: the experimenter announced to the class that someone had not picked up her reports at the last class session. He then said: "Will Miss Sandra Moore please come up and get her report." The young lady, a confederate[1] of the experimenter came up on the stage and stated in a loud, irritated tone of voice that one of her reports was missing. When the experimenter stated that she must not have handed it in, she angrily shouted: "I have

[1] The author wishes to express his appreciation to Christina B. Gidynski for her assistance in the classroom "drama."

never heard anything so ridiculous." (At this outburst, the class gasped and directed their full attention to the stage.) Then, "Sandra" pulled a water pistol from the pocket of her skirt, shot the experimenter with a stream of water, and fled from the stage dropping a piece of paper as she went.

Five minutes later when the class was restored to order, they were questioned about the incident. They wrote answers to the following questions:

What was the young lady's name?
What color was her blouse? Her skirt?
What color was her purse?
Did she drop anything?

Actually there was no purse. In asking about it we have used an "implicative question," one that assumes or suggests the presence of an element not actually present.

## RESULTS

All subjects were classified relative to the medians of the distributions of *Pt* scores and acquiescence (A) scores. Four classifications resulted: (*Pt*−, A+), (*Pt*−, A−), (*Pt*+, A−), and (*Pt*+, A+). For example, (*Pt*+, A−) would be the group of subjects above the median on the distribution of *Pt* scores and below the median on the distribution of A scores. The correlation between the *Pt* scores and the A scores in the present study was .34, $p < .01$ with 228 *df*. However, the variability of the scores was considerable with the result that the number of subjects in each group was almost equal.

The testimony of the subjects was scored for accuracy of report. The critical question was the implicative question. "What color was her purse?" This leading question influenced 34 subjects to indicate that there was a purse when in fact there was none. The remaining 196 subjects either left this question blank or wrote "none." The 34 influenced subjects did not differ from the other subjects in accuracy of response to the straight recall questions.

The central analysis for the present investigation was based on the distribution of the 34 influenced subjects in the two-variable personality classification. The frequency distribution shown in Table 1 produces a chi square of 11.05, $p < .02$ with 3 *df*. This result supports the hypothesis that persons both strong in acquiescent response set and low in score on the *Pt* scale are more subject to influence when giving testimony than individuals not showing this personality pattern.

TABLE 1
FREQUENCY WITH WHICH "INFLUENCED" AND "NONINFLUENCED" SUBJECTS BELONGED TO THE FOUR PERSONALITY PATTERNS

| | (*Pt*−, A−) | (*Pt*−, A+) | (*Pt*+, A−) | (*Pt*+, A+) |
|---|---|---|---|---|
| "Influenced" | 7 | 16 | 6 | 5 |
| "Noninfluenced" | 51 | 41 | 51 | 53 |

## DISCUSSION

A person usually would not be "led" in his testimony were he aware that an attempt was being made to influence him. A tendency to be acquiescent should also contribute to susceptibility to influence of this comparatively subtle type.

There is considerable evidence that scores on the *Pt* scale are related to the ability to demonstrate verbal awareness. Scores on the *Pt* scale have been shown to be inversely related to "repression" (Eriksen & Davids, 1955; Kuethe & Hulse, 1960). After verbal conditioning, high scorers on the *Pt* scale are more able to verbalize awareness of influence than are low scorers as was shown by Eriksen, Kuethe, and Sullivan (1958). The tendency of the high scorer on the *Pt* scale to consciously monitor his behavior has been discussed by Kuethe (1961).

Acquiescent response set might logically be expected to be related to the *Pt* scale as the latter contains mainly positively keyed items. This is, however, not the case as has been shown by De Soto and Kuethe (1959). The clear implication of the *Pt* items that "something is wrong" (this scale contains "obvious" rather than "subtle" items) overrides any tendency to agree with an item because of acquiescence. Also it should be noted that the persons of primary concern in this investigation are below the median on *Pt* and above the median on A. These subjects would of course not be found were there a perfect positive correlation between A and *Pt*. Such a situation is consistent with the view that scores on the *Pt* and similar scales are a disjunctive function of symptoms (real or imaginary) and the capability of claiming these symptoms on an inventory (De Soto & Kuethe, 1960).

The less than perfect correlation between any performance measure and scales like the *Pt* scale and the Manifest Anxiety Scale is not

entirely random error but is in part due to the fact that such scales are not unidimensional. When a scatter plot of *Pt* scores versus performance is examined, a triangular distribution is often noted (De Soto & Kuethe, 1960). The performance of subjects with low *Pt* scores is more variable than the performance of subjects with high *Pt* scores. This is to be expected if it is true that high scorers believe they have symptoms and have claimed them while low scorers may either lack the symptoms or have them but be unwilling or unable to claim them because of their defenses. High scorers are simply a more homogeneous group as far as personality dynamics are concerned.

It may well be that the A measure serves to identify those low scorers on the *Pt* scale who would have earned high scores had they not been prevented (consciously or unconsciously) by the implications of agreeing to the items on the scale. These subjects are using defenses related to repression or denial and as a result are not conscious of pressures from the environment and hence are likely to be influenced by a leading question. It is these subjects who do less conscious monitoring of their behavior and as a result show less verbal awareness of what has happened.

In contrast, those low scorers on the *Pt* scale who do not show acquiescent response set may be "true" low *Pt* subjects in the sense that while they are agreeing to all of the symptoms that they actually possess, they earn a low *Pt* score because they do not have many symptoms. These nondefensive subjects are more likely to be aware of influences from the environment and as a result to evaluate them critically.

Similarly, the high scorer on the *Pt* scale shows by the fact that he is capable of claiming symptoms on such an inventory that he is not using defenses of the repression or denial variety. He tends to consciously monitor his behavior and is perhaps, because of his higher anxiety, more likely to do so than are others. The tendency to monitor and maintain vigilance results in awareness of pressures from the environment which gives some immunity from influences such as leading questions and learning of the variety that requires that the individual not be aware that some aspect of his behavior is being reinforced.

A very critical point is the relation between *Pt* and acquiescence. While the high scorer on the *Pt* scale does show acquiescence, he does not earn his high *Pt* score because of it, in spite of the fact that most of the *Pt* items are keyed in the positive direction. The reliability of the *Pt* scale is not reduced when a semireversed scale is used. De Soto and Kuethe (1959) state:

> The intercorrelations between the oppositely keyed halves of the balanced *Pt* scale are so high, in fact, that acquiescence set can be ruled out as an important factor in *Ss*' performance on the scale.

When *Pt* scores are correlated with scores on other positively keyed scales a positive correlation is found (Couch & Keniston, 1960). It is easy to assume, as did Couch and Keniston, that the correlations exist because acquiescence contributed to scores on all the scales. This is not necessarily the case. High scorers on the *Pt* scale earn high scores on positively keyed scales because acquiescence is part of their personality pattern, not because the items on the *Pt* scale are also positively keyed. High scorers on the *Pt* scale will deny the absence of symptoms just as readily as they will agree to the presence of symptoms; their behavior is essentially "symptom claiming" (De Soto & Kuethe, 1959).

There is a distinct possibility that the acquiescent response set may function in the way that the *K* scale of the MMPI was intended to function but does not, at least with respect to the *Pt* scale. The *K* scale cannot effectively "correct" the *Pt* scale for defensiveness. Fordyce (1956) reports a correlation of −.75 between the *K* and the *Pt* scales. The reliability of the *K* scale is .76 (Hathaway & McKinley, 1951); the *K* scale could not, therefore, show a closer relation to the *Pt* scale than it does. These two scales may be regarded as measures of the same thing, with the *Pt* scale being the better measure because it is more reliable. Study of the item content of these two scales reveals that many of the items refer to similar behavior and symptoms except that they are keyed in opposite directions, therefore resulting in the negative correlation. On both the *Pt* scale and the *K* scale a person has the same task: accepting or rejecting items that inquire about personal behavior and feelings. This view is supported by the correlations reported by Edwards (1957); the correlation between

the Social Desirability Scale (SDS) and the *K* scale is .81 while the SDS and the *Pt* scale correlate −.85. As the split-half reliability of the SDS is .83, once again the picture is that of scales intercorrelating to the limits of their reliabilities.

In the present study, acquiescent response set has demonstrated its usefulness in identifying two distinct behavior patterns among low scorers on the *Pt* scale. This result supports the possibility of "suppressor variable" functions for the acquiescent response set, which has earned the right to be regarded as a personality variable and a very interesting one at that. Among its virtues it can be easily measured and with high reliability (e.g., a split-half *r*, without correction for test length, of .70 was reported by Kuethe, 1959. However, the present analysis indicates that it would be wrong to go to the other extreme and regard acquiescence as that which is being measured on most of the standard personality inventories. There is far more to an individual's personality than acquiescence; the important problem is the exploration of the ways in which this variable interacts with other personality variables to produce behavior variations from individual to individual and from situation to situation. This is, obviously, an area in which great caution is required; the investigation of a variable that does influence so much of behavior encounters unique problems. As was shown in the foregoing discussion, it is possible for a person to earn a high score on a positively keyed personality inventory, show acquiescence in a variety of situations, and yet have earned the high score on the inventory for reasons other than acquiescence. It would be easy to overlook this possibility and assume that acquiescence produced the high score on the positively keyed inventory—a case of the classic logical error of misattributing the direction of cause and effect between two correlated variables. It appears in the present situation that a person does not get a high score on the *Pt* scale because he is acquiescent, but rather, he is acquiescent because this is part of the personality pattern of the individual who earns a high score on the *Pt* scale.

## SUMMARY

The classical testimony experiment was used to explore the relation between the acquiescence response set and a personality measure, the Psychasthenia scale of the MMPI. It was hypothesized that subjects who could be "led" in their testimony would have a strong acquiescence response set and be low scorers on the Psychasthenia scale. This hypothesis was supported by the testimony of 230 subjects who had witnessed a classroom "drama."

Evidence was discussed that the high scorer on the *Pt* scale is "acquiescent" but does not earn his high score on the positively keyed *Pt* scale because of this acquiescence. A "suppressor variable" function for the acquiescence response set was suggested.

## REFERENCES

BURTT, H. E., & GASKILL, H. V. Suggestibility and the form of a question. *J. appl. Psychol.*, 1932, **16**, 358–373.

COUCH, A., & KENISTON, K. Yeasayers and naysayers: Agreeing response set as a personality variable. *J. abnorm. soc. Psychol.*, 1960, **60**, 151–174.

DE SOTO, C. B., & KUETHE, J. L. The set to claim undesirable symptoms in personality inventories. *J. consult. Psychol.*, 1959, **23**, 496–500.

DE SOTO, C. B., & KUETHE, J. L. On the relation between two variables. *Educ. psychol. Measmt.* 1960, **20**, 743–749.

EDWARDS, A. L. *The social desirability variable in personality assessment and research.* New York: Dryden, 1957.

ERIKSEN, C. W., & DAVIDS, A. The meaning and clinical validity of the Taylor anxiety scale and the hysteria-psychasthenia scales from the MMPI. *J. abnorm. soc. Psychol.*, 1955, **50**, 135–137.

ERIKSEN, C. W., KUETHE, J. L., & SULLIVAN, D. F. Some personality correlates of learning without verbal awareness. *J. Pers.*, 1958, **26**, 216–228.

FORDYCE, W. E. Social desirability in the MMPI. *J. consult. Psychol.*, 1956, **20**, 171–175.

HATHAWAY, S. R., & MCKINLEY, J. C. *Manual for the Minnesota Multiphasic Personality Inventory.* (Rev. ed.) New York: Psychological Corp., 1951.

KUETHE, J. L. The positive response set as related to task performance. *J. Pers.*, 1959, **27**, 89–94.

KUETHE, J. L. The interaction of personality and muscle tension in producing agreement on commonality of verbal associations. *J. abnorm. soc. Psychol.*, 1961, in press.

KUETHE, J. L., & HULSE, S. H. Pessimism as a determinant of the tendency to claim undesirable symptoms on personality inventories. *Psychol. Rep.*, 1960, **7**, 435–438.

MUSCIO, B. The influence of the form of a question. *Brit. J. Psychol.*, 1916, **8**, 351–389.

WHIPPLE, G. M. The observer as reporter: A survey of the "psychology of testimony." *Psychol. Bull.*, 1909, **6**, 153–170.

(Early publication received October 12, 1960)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 323-328

# AFFECT-STIMULATING EFFECTS OF COLORS[1]

ROBERT J. DRECHSLER[2]
*Yale University*

THE present study tested two hypotheses: colored stimuli elicit more emotional responses than do gray stimuli; different colors have different quantitative effects in this regard, with red the most stimulating. Although these contentions underlie much of the writing and clinical practice involving the Rorschach (1942) test, they have rarely been closely and systematically studied.

The controversy surrounding the alleged affect-stimulating characteristics of chromatic stimuli may be found in articles by Schachtel (1943), Rapaport, Gill, and Schafer (1946), Rickers-Ovsiankina (1943), Klopfer and Kelley (1942), Lazarus (1949), Siipola (1950), and Siipola, Kuhns, and Taylor (1952). Attempts to test the validity of the color-affect relationship have suffered from dependence on the Rorschach cards as crucial materials for testing the hypotheses. In very few cases (Haward, 1955; Holzberg, 1955) was the color variable removed from the specific context of the Rorschach cards. In every case, the criterion for emotional involvement of the subjects (*S*s) has been some measure of the way in which they solved problems *in which color played an integral part*. Such methods justify the criticism offered by Siipola (1950). As long as color is one element to be used in solving a problem, the differences observed between performances with colored versus achromatic stimuli may rest on differences in the difficulty of the two problems rather than on the emotionally stimulating properties of color. The present experiment attempted to overcome this methodological difficulty.

## METHOD

The general technique was to show homogeneous rectangles of bright red, green, and gray on a screen, and to follow each presentation of a projected stimulus with an orally presented word-association test. The colors were presented under two conditions: continually through the word-association test, and in a 5-second exposure followed immediately by a word-association test. The use of long and short color exposures permitted evaluation of the disruptive effect colors may have while a task is being performed as compared with their continuing effect once they are removed.

For half the test exposures, the visual stimuli were exposed for 5 seconds, removed, and followed by an orally-presented five-item word-association list; for the other half the exposures, the visual stimuli were presented continuously through the oral presentation of a five-item list of words. In the latter instance, the exposure of each color began 5 seconds before the presentation of the first word in order to eliminate any disruption of response due to an abrupt presentation of the stimulus. The procedure also entailed a period of stimulation prior to word association that was identical with that for the noncontinuous color condition. One color (red, green, or gray) was exposed with each five-item word list. There were six stimulus conditions: red continuous exposure, red short exposure, green continuous exposure, green short exposure, gray continuous exposure, and gray short exposure. Each of the six conditions was presented four times to each *S*; in all, 20 words were presented with each stimulus condition, five at each presentation. There were, therefore, 24 five-item word lists, making a total of 120 words, which were taken from the Orbison (Rapaport et al., 1946, Vol. II, p. 13) and Kent-Rosanoff (Rosanoff, 1938) word lists for which relevant norms of responses are available (Lesser, 1948; Russell & Jenkins, 1954).

Each *S* served as his own control. The order of presentation of the six stimulus conditions was such that for each *S* each condition followed every other condition, except that no one color was ever shown on two consecutive presentations.

Between *S*s, the order of presentation was systematically varied, so that over the entire experiment, each of the six conditions occurred four times with each word list and in each sequential position. The 24 word lists were always presented in the same sequence; but within lists, the order of presentation of the individual words was systematically varied to control for possible effects due to position within the series.

A recall test for the entire list of 120 words was run for each *S* after all the association tests and stimulus conditions had been completed. Reaction times to the stimulus words for both the association and recall sections of the experiment were measured to .01 second with an electric timer. The timer was set in motion by the experimenter (*E*) when he finished pronouncing each word. It was stopped when he heard *S*'s voice.

*S*s were seen individually and were tested for red-green color blindness before the experiment. No *S* was

[1] This paper is adapted from a dissertation submitted to Yale University in 1956 in partial fulfillment of the requirements for the PhD degree. The author wishes to express his deep appreciation to Seymour B. Sarason, Robert P. Abelson, and Roy Schafer for their many constructive suggestions in the design and execution of this investigation.

[2] Now at Pittsburgh Child Guidance Center.

color blind. The test room was dark except for a small shaded lamp on *E*'s desk. *Ss* faced a screen from a distance of about 8 feet. The projected colored rectangles were 45″ × 32″. The word association instructions used by Rapaport et al. (1946), as adapted to the present purpose, were used.

All *Ss* were male students in elementary psychology courses at Yale University. Selected at random from a list of students in the courses, they ranged in age from 18 to 23, the median and modal age being 19.

## RESULTS

In terms of the experimental design, two primary hypotheses were tested: there are more word-association disturbances under both red and green conditions than under gray conditions; and red conditions elicit more word-association disturbances than will green conditions.

In terms of relatively long reaction time (RT) to stimulus words as the index of disturbance, the results yielded highly significant differences ($p < .01$)[1] between colors. Responses under the red conditions had the longest RTs; responses under the gray conditions had the shortest. When the two degrees of freedom between colors are divided into Red-plus-Green versus Gray with 1 *df*, and the orthogonal comparison of Red versus Green with 1 *df*, both of these comparisons are significant at the .05 level of significance. The magnitude of differences between color groups was small but consistent. The average RT was 1.40″ for the red conditions, 1.31″ for the green conditions, and 1.27″ for the gray conditions. Of the 24 *Ss*, 16 had their longest mean RT under red conditions, 5 under green, and 3 under gray. The chi square test (with Yates correction) of the deviation of this distribution from chance expectancy was significant at the .01 level. These results support Hypotheses 1 and 2.

Although over-all analysis of variance showed no significant differences between continuous and short exposures of colors, the data were further analyzed to determine whether the obtained results held equally for the continuous and short exposure times. The following additional analyses of variance were computed:

*Performance RT—Continuous exposure only.* Between colors was significant with $p$ less than .01. For the comparison Red-plus-Green versus Gray, $p$ fell between .05 and .10. For the comparison Red versus Green, $p$ was less than .05.

*Performance RT—5″ exposures only.* Between colors was not significant, nor were Red-plus-Green versus Gray or Red versus Green. These results indicate that the obtained significant differences depend primarily on the long exposure conditions.

*Relatively long reaction times in recalling the original word association responses.* The analysis of variance between colors was significant with $p$ less than .01. The comparison Red-plus-Green versus Gray was significant with $p$ less than .01. Red versus Green was significant at the .05 level. The mean recall RTs were 1.32″ for red, 1.22″ for green, and 1.16″ for gray. Eighteen *Ss* had their longest mean RT for red, 4 for green, and 2 for gray. A chi square test (with Yates correction) of deviation of this distribution from chance expectancy was significant at the .01 level. These results support the hypotheses.

As was the case with performance RTs, over-all analysis of variance showed no evidence of significant differences between continuous and short exposures of colors. The data were further analyzed as follows:

*Recall RT—Continuous exposures only.* Between colors was significant with $p$ less than .01. Red-plus-Green versus Gray was significant with $p$ less than .01. Red versus Green was not significant.

*Recall RT—5″ exposures only.* Between colors was not significant. Neither Red-plus-Green versus Gray nor Red versus Green was significant. These results suggest that the continuous exposure conditions contributed more than the short exposure conditions to the over-all difference between colors and to the distinction between the two chromatic colors and gray.

An over-all analysis of variance of incorrect recalls of initial responses, taken as a disturbance index, did not yield significant differences between colors. However, examination of the data indicated that although the means for the different color conditions were almost identical for the short exposure time conditions, the means for the continuous exposure time conditions showed systematic differences. Therefore, this latter half of the data was analyzed separately. Differences between means for the different color conditions were then found to be significant with $p$ less than .05. Red-plus-Green versus Gray was not significant, although the differences were in the predicted direction. Red versus Green was significant with $p$ less than .05. This finding supports Hypothesis 2 and is consistent with Hypothesis 1, but only for the continuous color exposure conditions.

Unusualness of response word, based on

[1] All $p$ values are based on a two-tailed distribution.

norms from 1,000 university students for the Kent-Rosanoff word list (Russell, 1954) and on Lesser's (1948) norms from 400 New York City *Ss*, not exclusively university students, for the Orbison word list, was also taken as an index of disturbance. For each stimulus word, the response words were listed in the order of frequency with which they were given in the normative group. With the responses arranged in descending order of popularity, cut-off points for scoring were established. Those responses which included 50% of the normative population (starting with the most popular response and adding frequencies down the list) were given an Unusualness score of 0; those responses which included the range from 50% to 75% of the normative population were given a score of 1; the least popular responses given by the remaining 25% of the population were scored as 2. Responses unique to the present experimental group were also given a score of 2. Analyses of variance were computed for the distribution of responses with weighted scores of 1 and 2 and for the Score 2–responses alone. These two analyses yielded practically identical results. Over-all analysis of variance showed no evidence of significant differences between color conditions. However, when the 2 degrees of freedom for the Between Colors measure were broken down into Red-plus-Green versus Gray and Red versus Green, the former measure was significant at the .05 level. These results support Hypothesis 1.

In addition, a score was devised based on the responses of the 24 *Ss* in the experiment. Each response that occurred only once was considered "unique" and was given a weighted score. The weight was the reciprocal of the total number of different responses to the particular stimulus word in this group of *Ss*. The weights represented an attempt to reflect arithmetically the idea that a unique response to a word that ordinarily yields but few different responses rates a higher "disturbance" score than a unique response to a stimulus word that elicits a wide variety of responses. Analysis of variance showed no significant differences between colors. Although the rank order of red, green, gray did not hold for this measure, red and green both elicited more unique responses than did gray. The results were not statistically significant, however.

Finally, unusual response words were rated as disturbed or not disturbed by a psychologist[4] who was supplied with a list of all responses to all stimulus words. Those responses which had received a 0 score for Unusualness were omitted from consideration. Also omitted were all responses which had a frequency of 25 or more on the Russell and Jenkins (1954) norms for the Kent-Rosanoff list, or a frequency of 10 or more on the Lesser (1948) norms for the Orbison list. The rater did not know the experimental condition under which a response was obtained, nor did he see the pattern of responses given by any individual *S*. Analysis of variance for this measure showed no significant differences between groups; hence no detailed analysis or reliability study of this qualitative measure was attempted. However this measure, in common with all other measures, showed gray conditions to elicit the fewest disturbed responses.

Analysis of variance for each of the six disturbance measures also tested differences between *Ss* and the various interactions of colors, *Ss*, and durations of exposure. For every disturbance measure, the differences between *Ss* were highly significant with $p$ less than .001. On the performance RT measure, the interaction of Colors × Subjects approached significance at the .05 level. This interaction, under continuous exposure conditions only, was significant at the .01 level. Different *Ss* thus reacted differently to the different colors despite the fact that the preponderant tendency was for the rank order of disturbance scores on this measure to be Red, Green, and Gray. No other main comparison or interaction approached statistical significance on any of the disturbance measures.

After each disturbance measure had been analyzed, rank-order correlations of performance were computed, a rank being assigned to each *S* according to the magnitude of his contribution to the obtained differences between the color groups. In this way a measure was obtained of tendency for differences found on different measures to stem from the same *Ss*. In each measure, the total score for each *S* was used to ascertain his rank. Table 1 shows that there was a significant tendency for the

[4] The writer thanks Roy Schafer for making these ratings.

TABLE 1
RANK-ORDER CORRELATIONS AMONG THE SIX TEST MEASURES

| | RT(P) | RT(R) | IR(C) | IR(T) |
|---|---|---|---|---|
| RT(P) | | | | |
| RT(R) | .49* | | | |
| IR(C) | .25 | .26 | | |
| IR(T) | .23 | .26 | .55** | |
| UR | .12 | .30 | .24 | .44* |

Note.—Abbreviations:
RT(P)—Reaction time-performance
RT(R)—Reaction time-recall
IR(C)—Incorrect recall-continuous exposure only
IR(T)—Incorrect recall-total scores
UR—Unusualness of response words.
* $p < .05$, two-tailed test.
** $p < .01$, two-tailed test.

same *S*s to contribute to disturbance measures on both performance and recall RT and on unusualness of response and incorrect recall. The significant correlation between IR(C) and IR(T) is an artifact; IR(C) is contained within IR(T).

Further analysis of disturbance measures for each *S* was performed to determine whether the different types of disturbance shown by each *S* tended to cluster together in reaction to specific words and specific color conditions. On single stimulus words, scores for each *S* were dichotomized as scored plus or minus, depending on whether the response was "disturbed" or not. Chi square tests were performed between the measures for each *S* and summed for all *S*s. Summing the individual chi squares was justified by the fact that for all comparisons and all *S*s the predominant tendency was for both measures to be in the same direction, i.e., both disturbed or both not disturbed. For RT measures, dichotomization was accomplished by dividing the RTs at the median, with medians for each *S* computed separately for his red, green, and gray conditions. RTs that were longer than the median RT (for *S* on each color condition) were scored plus; those below the median were scored minus. For other measures, each response was scored "disturbed" or "not disturbed." Only measures that had shown significant results on analysis of variance were compared in this way. The measures paired in this way included RT (Performance) versus RT (Recall), RT (Performance) versus Incorrect Recall, RT (Recall) versus Incorrect Recall, RT (Performance) versus Unusualness of response, RT (Recall) versus Unusualness of response, and Unusualness of response versus Incorrect Recall. For every pair of measures, the summed chi squares were significant at the .01 level or better.

## DISCUSSION

The results generally confirm Hypotheses 1 and 2. All the analyses give results consistent with Hypothesis 1; analyses of RT data for both performance and recall, as well as of the Incorrect Recall data for the continuous exposure condition, support Hypothesis 2. The most striking differences between color conditions were obtained when the visual stimuli were present before and during the time the word association lists were presented. This finding suggests that with a normal population there is rapid recovery from the emotionally stimulating effects of colors as soon as the colors are removed. The direct relation to the color variable of length of RT and occurrence of unusual responses on the performance section of the present experiment appears to be established.

A question may be raised, however, as to whether recall disturbances bear a similar direct relation to the color variable. The disturbance measures clearly were not independent of each other. There was a significant tendency for stimulus words to elicit responses which were rated "disturbed" on more than one measure, and this tendency was demonstrated for every pair of measures. During the recall part of the test the colored slides were not shown. Recall RT and incorrect recall of initial responses were both highly related, within *S*s, to performance RT. These data suggest that recall disturbances may have depended on the reactions occurring during the performance section rather than on a delayed effect of the colored stimuli themselves.

The significant correlation between *S*s of unusualness of response and incorrect recall indicates that an unusual response is harder to recall than a usual one. This fact alone would account for increased RT on recall for unusual words. *S* tries to remember the initial response, sometimes fails to do so, and finally gives an incorrect response. In this way the high correlation between recall RT and incorrect recall can be accounted for without re-

course to the color variable. The crucial disturbance here is probably the unavailability of the original response; it is the test instructions that shift the emphasis to recall RT and incorrect recall.

All the disturbance measures used in this study have previously found some validation as emotional indicators. The validation has come primarily from comparisons of diagnostic groups with normals and of emotionally loaded stimulus words with neutral words (Hull & Lugoff, 1921; Jung, 1919; Luria, 1932; Rapaport et al., 1946; Rosanoff, 1938; Schumer, 1949). In the present experiment, these measures were used with a relatively homogeneous normal population and with a word list from which most of the "critical" words had been removed.[5] It is not surprising, therefore, that the magnitude of the differences between groups should be much less for the present experiment than for the groups on which validation of the word association test was based. The purpose of neutralizing the word list and of using normal *S*s was to minimize the variance in disturbance scores due to factors other than the color variable.

Rapaport et al. (1946) have elaborated the idea that evidence of selective disturbance in the word association test generally points to the stimulus word as the source of disturbance. In their findings, the specific response word usually did not reveal the nature of the disturbance, but only that some disturbance had occurred. In the present experiment, stimulus conditions must be looked to for the source of disturbance. The consistency with which disturbance differences occurred between color conditions is evidence for this.[6]

It may be argued that *E*, with full knowledge of the hypotheses, may unwittingly have influenced such a delicate measure as RT in the desired direction. That this was not the cause of the differences between color conditions is indicated by the following considerations:

1. RT was not the only measure to show systematic differences between colors.
2. Within-subject chi squares between RT and other disturbance measures were consistently significant, indicating that stimulus words which elicited long RTs also elicited other signs of disturbance.
3. Mean differences between RTs for the different colors were of the magnitude of .10 second, which would have required a most sensitive preconscious mechanism to produce.

Results of this experiment challenge Siipola's (1950) hypothesis that color itself has no emotional impact and that difficulty in handling Rorschach color cards is due only to the increased difficulty of the problem presented when color has to be integrated with form. In Siipola's formulation, emotional accompaniments of color responses stem from the *S*s' ego involvement in a difficult problem, not from intrinsic properties of colors. The present experiment did not require *S*s to integrate color into their responses. If the differences between colors and gray were due to *S*s' tacit assumption that the visual stimuli should be incorporated in their responses, it is unclear why chromatic stimuli should have presented more of a problem than gray because the visual stimuli were homogeneous and rectangular, and the stimulus words had no particular reference to red, green, or gray.

A possible explanation for the role of colors as affect-stimulating agents is that bright colors intrude on the perceiving person in such a way that his ongoing thought processes are interrupted or disrupted until the intrusive stimulus can be removed or otherwise assimilated. This conception of the intrusive nature of color has been elaborated by Schachtel (1943). The intrusion of color is similar to the effects of more specifically threatening environmental or internal stimuli, and for this reason it may either exaggerate strong defenses that have been developed against such threats or disrupt these defenses if they are weak.

The prepotent responses to different colors may themselves be disturbing in different degrees. For example, the common associations of "blood," "fire," "anger," and "passion"

[5] In selecting 120 words for which relevant normative data existed, it was impossible to eliminate all potentially critical words. Possibly any such word list, even without restrictions, would contain critical words. In the present list the words "love," "bite," "health," and "hungry" clearly pertain to common areas of conflict.

[6] A significant interaction of Colors × Stimulus words could account for the results, although words obviously related to any of the visual stimuli were not included on the list. This interaction was not tested, and in any case could not have invalidated the general findings that red elicited more disturbances than green, and both red and green elicited more disturbances than gray.

with red, and the disturbing connotations of these associations may account for the consistency with which red stimuli elicit the longest RTs. The significant interaction of Colors × Subjects under continuous exposure conditions for performance RT indicates that different *S*s react differently to different colors. This argues against the universal efficacy of one color over another or over gray, and it is consistent with the idea that individual associations to a particular color or individual differences in type and efficacy of defense measures may determine the extent to which that color produces measurable disturbances in word association responses.

The conception of color disturbance presented here is not intended to contribute to the *mystique* surrounding some interpretations of Rorschach color shock. It is not suggested that color is the only variable that may produce the disturbances observed. Any sharply intrusive variable would be expected to have a similar effect.

## SUMMARY

It was the purpose of this experiment to test the following hypotheses: colored stimuli elicit more emotional responses than do gray stimuli; different colors have different quantitative effects in this regard, with red being the most disturbing color. While these hypotheses are central to Rorschach test interpretation, they have not previously found clear-cut experimental support.

The method used was the projection on a screen of homogeneous rectangles of bright red, green, and gray. Each presentation of a visual stimulus was followed by an orally-presented word-association test. Several measures of disturbance on this test were used to measure disturbance elicited by the visual stimuli.

Several disturbance measures revealed statistically significant differences between colors and all supported Hypothesis 1. Reaction time measures showed significant differences in support of Hypothesis 2. The measures of disturbance proved to be interdependent, a fact which created some ambiguity with regard to details of interpretation, but did not interfere with the major conclusion that the hypotheses were supported.

The results were tentatively interpreted in terms of an "intrusion" hypothesis. Colors were thought to intrude on the perceiver in a manner similar to the intrusions of more specifically threatening environmental or internal stimuli and to arouse or break through the defenses used against threats. The degree of disturbance caused by a particular color was thought to depend on the disturbing connotations of the associations it arouses or on the type and efficacy of defense stimulated thereby.

## REFERENCES

BECK, S. J. *Rorschach's test.* Vol. I. New York: Grune & Stratton, 1944.

HAWARD, L. R. C. Colour associations as a contributing factor to neurotic colour shock. *Brit. J. med. Psychol.*, 1955, **28**, 183–187.

HOLZBERG, J. D., & SCHLEIFER, M. J. An experimental test of the Rorschach assumption of the impact of color on the perceptual and associative processes. *J. proj. Tech.*, 1955, **19**, 130–137.

HULL, C. L., & LUGOFF, L. S. Complex signs in diagnostic free association. *J. exp. Psychol.*, 1921, **4**, 111–136.

JUNG, C. G. *Studies in word association.* New York: Moffat, Yard, 1919.

KLOPFER, B., & KELLEY, D. *The Rorschach technique.* Yonkers: World Book, 1942.

LAZARUS, R. S. The influence of color on the protocol of the Rorschach test. *J. abnorm. soc. Psychol.*, 1949, **44**, 506–516.

LESSER, G. S. Commonality of response to the Orbison Word-Association Test. Unpublished master's thesis, Columbia University, 1948.

LURIA, A. R. *The nature of human conflicts or emotion, conflict and will.* (Trans. by W. H. Gantt) New York: Liveright, 1932.

RAPAPORT, D., GILL, M., & SCHAFER, R. *Diagnostic psychological testing.* Vol. II. Chicago: Year Book Publishers, 1946.

RICKERS-OVSIANKINA, M. Some theoretical considerations regarding the Rorschach method. *Rorschach res. Exch.*, 1943, **7**, 41–53.

RORSCHACH, H. *Psychodiagnostics: A diagnostic study based on perception.* (Trans. by P. Lemkau & B. Kronenberg) New York: Grune & Stratton, 1942.

ROSANOFF, A. J. *Manual of psychiatry.* New York: Wiley, 1938.

RUSSELL, W. A., & JENKINS, J. J. *The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test.* Minneapolis: Univer. Minnesota, 1954.

SCHACHTEL, E. G. On color and affect. *Psychiatry*, 1943, **6**, 393–409.

SCHUMER, F. C. Some behavioral correlates of Rorschach human movement response. Unpublished doctoral dissertation, Yale University, 1949.

SIIPOLA, ELSIE M. The influence of color on reactions to ink blots. *J. Pers.*, 1950, **18**, 358–382.

SIIPOLA, ELSIE, KUHNS, F., & TAYLOR, V. Measurement of the individual's reactions to color ink blots. *J. Pers.*, 1952, **21**, 22–47.

(Received May 11, 1959)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 329-334

# AUTHORITARIANISM AND MEN'S REACTIONS TO SEXUALITY AND AFFECTION IN WOMEN[1]

RALPH ROTHSTEIN

*Michael Reese Hospital, Chicago, Illinois*

THIS paper was stimulated by Gollin's studies (1951, 1954) on the formation of impressions of personality. In these experiments male subjects (*Ss*) were shown a movie consisting of scenes in which the same woman portrayed two diverse roles, that of the prostitute and the kindly woman. Gollin found that impressions of personality were formed along three distinct lines. Some *Ss* described the actress in terms of only one character quality and ignored the other (simplified or unitary impressions). Others characterized her in terms of both behavioral themes but did not relate them (aggregated impressions). There were still others who, in addition to noting the divergent behavioral patterns, integrated them in a meaningful and consistent manner (integrated or related impressions). Gollin has proposed that such differences in impression formation can be accounted for in terms of "cognitive-organizing" processes in the perceiver (Gollin, 1954, 1958; Gollin & Rosenberg, 1956) and has begun to explore the possible sources of these differences. In his recent experiment with children he found that age, sex, intelligence, and social background were all significant factors.

Authoritarianism should also be related to different cognitive modes of organizing impressions of others. If we assume that the problem of forming an impression of someone who displays inconsistent or disparate behavioral tendencies represents an ambiguous perceptual task, it follows that authoritarian persons, because of their presumed intolerance of ambiguity, should tend to form nonintegrated impressions. This formulation seems applicable regardless of the specific content of the judged person's inconsistencies. Gollin was apparently not concerned with the specific role of content in organizational patterns, and did not discuss why he selected promiscuity and kindness in a woman as the diverse character qualities, in preference to other discrepant characteristics, e.g., stinginess vs. generosity. However, in the present study, the specific behavioral characteristics (sexuality and kindness) displayed by the stimulus person were regarded as relevant, apart from the formal inconsistency of behavior, as groups differing in authoritarianism were expected to react differentially to the sexual and kindly cues provided by the stimulus person.

Among the variables found by Adorno, Frenkel-Brunswik, Levinson, and Sanford (1950) to be associated with authoritarianism was an exaggerated concern with sexuality, which the authors suggest, reflects a tendency to project unacceptable sexual impulses on to others. Frenkel-Brunswik's clinical interviews further showed male authoritarians to possess "dichotomous sex attitudes"; that is, they separated, or were not able to integrate, sexuality and affection in feeling and behavior.

Authoritarianism was therefore considered to be a personality characteristic of the perceiver significant to the process of social perception, especially when the stimulus person's behavior connotes sexuality and kindness (affection). While the importance of authoritarianism as a variable in impression formation has been demonstrated in several investigations (e.g., Jones, 1954; Scodel & Freeman, 1956; Scodel & Mussen, 1953; Thibaut & Riecken, 1955), these studies have dealt with aspects of behavior relating to leadership, insightfulness, aggression, etc., rather than with the role of authoritarianism in men's social perception of women.

On the basis of the findings emerging from *The Authoritarian Personality*, the following hypotheses were framed:

1. Authoritarians should tend to overemphasize the sexual aspects of a woman's behavior, even though she is also shown to be affectionate and kindly as well.

2. Authoritarians should show dichotomous

[1] This research is based in part on a dissertation submitted in partial fulfillment of the requirements for the PhD degree, University of Massachusetts, 1959.

The author wishes to thank Solis L. Kates, Seymour Epstein, and Claude C. Neet for their helpful criticisms and suggestions in carrying out the investigation.

sex attitudes as evidenced by their "splitting" of sexuality and affection. Further, their impressions of the personality of a woman who is both sexual and kindly should tend to be organized in a nonintegrated manner.

## METHOD

### *Subjects*

One hundred and twenty male undergraduates enrolled in the introductory psychology course at the University of Massachusetts were used as *Ss*. Of this group 32 were selected as high authoritarians and 32 as low authoritarians.[2]

### *Stimulus Material*

The stimulus situation consisted of a silent motion picture containing two scenes. One person, a young, attractive female in her middle twenties, appeared in both scenes and henceforth will be referred to as the star. In Scene 1 the star portrayed the "sexy" female, whose physical appearance and behavior were intended to characterize her as sexually appealing. Although the star was seen as sexually attractive, her behavior was not intended to suggest sexual immorality or promiscuity. In Scene 2, the star's behavior characterized her as being kind, affectionate, and understanding. An attempt was made to avoid any behavior that might be interpreted as sexually suggestive. It was intended that the star's behavior be equally convincing in both scenes, so that "objectively" one could describe her as both sexual and kindly to about the same degree. Each scene was approximately 5 minutes in length. The movie was presented to three other psychologists for judgment and there was agreement that the kindly cues were at a minimum in the sexual scene and sexual cues at a minimum in the kindly scene. A brief description of each scene is presented below:

*Sexual Scene.* As the scene opens, a young woman in a black slip is sprawled out on a bed. Her eyes are closed and her hair is somewhat disheveled. After a few seconds, she gets out of bed, yawns lazily and proceeds to groom and dress herself. Her mannerisms are sensually suggestive (e.g., she admiringly passes her hands over her body, carefully strokes her hair, etc.). She puts on her stockings, paying close attention to whether the seams are straight. She is shown attaching the stockings to her girdle, so that her thighs are fully revealed. At one point in this sequence, the camera is raised above her head and focuses on her breasts, which at this angle are partially exposed. She then puts on a tight fitting, low-cut dress and is shown carefully applying make-up. As the scene closes she is shown embracing her date, who has called for her; her general attitude is one of indifference.

*Kindly Scene.* As the scene opens, the star is reading a book. She is wearing a loose fitting jumper and blouse, so that her bodily attractiveness is minimized. On the table close to her is the photograph of a young man in army uniform. As she looks at it, a tender, affectionate expression comes over her face. After getting some stationery, she commences to write a letter, the content of which clearly reveals her love for the man in the photograph. The following is an excerpt from the letter which was, in its entirety, presented to *Ss*:

> Although you never complain in any of your letters, I know that things are pretty rough out there. I only wish that there was something I could do to help to make you feel the warmth of home and to show you how much I love you. Please, darling, take care of yourself; you mean everything to me.

As she finishes the letter, a boy of about 13 enters the room and is greeted by the star, who messes his hair in an affectionate gesture. He sits down with a magazine and the star leaves the room, returning shortly with a snack for him. After eating, the boy begins his school work. He has difficulty with his lesson and goes to the star, who helps and encourages him with his work.

### *Procedure*

The California F Scale (modified Form 45) was administered to all 120 *Ss*. Of the 30 items on the F Scale employed in this study, 26 were the same as those in Form 45, with the other 4 items only slightly modified. The mean F score for the high and low authoritarian groups, each composed of 32 *Ss*, was 130.25 and 101.63, respectively.[3] Two weeks after the F scale was administered, *Ss* were shown a silent movie in groups of no more than 12. Two treatment conditions were used in presenting the movie. Half the group was shown the sexual scene first (SK presentation) and for the remaining half, the kindly scene was presented first (KS presentation). After the movie was shown, *Ss* were asked to write their impressions of the personality of the woman who appeared in the movie.

### *Written Impressions*

Gollin's (1951) criteria were used in categorizing the impressions into type of organization employed. In brief, these were: (*a*) unitary—in which one major

[2] The 64 *Ss* who participated in this study were originally selected on the basis of their performance on a specially constructed set of TAT cards depicting mother-child interactions. An attempt was made to differentiate *Ss* on the basis of approach-avoidance conflict over sexual feelings toward their mothers. This measure, however, was only partially successful. *Ss* were significantly differentiated only on the basis of the avoidance measure. The specific hypotheses which were then formulated for this "conflict" group were not substantiated. A complete discussion of this aspect of the research can be found in the author's doctoral dissertation, Men's Perceptions of Sexuality and Kindness in a Woman, University of Massachusetts, 1959.

[3] Only the F Scales of those 64 *Ss* already selected to serve in the experiment on the basis of their performance on the TAT measure of "avoidance sexual conflict" toward their mothers were scored. In both authoritarian groups, there was an equal number of *Ss* who were considered high and low in avoidance conflict over feelings toward mother.

theme (the star's sexuality or kindness) dominates the impression; (*b*) aggregated—in which the two major themes are present, but treated as independent and unrelated aspects; (*c*) integrated—in which both themes are present, each being related to the other, while the independence of each is maintained. To study the overall affective tone of the written impressions, each protocol was classified according to whether it was favorable, unfavorable, or both favorable and unfavorable. Organizational type and affective tone were judged independently of each other. In classifying the impressions, numbers were substituted for *S*s' names and all other identifying information was concealed.

### *Movie Rating Scale*

To complement the free written description, a rating scale was administered after the written impressions of the star had been collected. The rating scale contained 30 randomly arranged items, 15 relating to sexual behavior, and 15 to kindly and affectionate behavior, each of which *S* was to mark according to one of four categories: strongly agree (+2), agree (+1), disagree (−1), and strongly disagree (−2). Examples follow, the first three sexual and the latter three kindly:[4]

If she were to buy a bathing suit, she would probably select one that would be revealing and tight fitting.

She has a great deal of flashy sex appeal.

The way she walks and dresses probably attracts men's interest in her.

She probably would enjoy doing charity work in which she could help others.

She appears to be a sweet, understanding young woman.

She appears to be the type of woman who, if she quarreled with her husband, would probably feel badly and want to make up.

Prior to its use in the study proper, the items of the scale were administered to 10 male graduate students who were asked to indicate whether an item indicated kindness or sexuality in a woman. Only the items with at least 90% agreement were retained. On all items but one, judges were unanimous.

### *Heterosexual Questionnaire*

A questionnaire of 12 items indicating splitting of sexuality and affection in heterosexual behavior was constructed by the experimenter (*E*) and submitted to three other psychologists for judgment. The final form of the questionnaire consisted of the eight true-false statements that were considered most appropriate. These also allowed four responses: definitely true (+4); mostly true, but not entirely so (+3); mostly false, but not entirely so (+2); and definitely false (+1). The questionnaire was presented after the movie rating scale was completed, with strong assurances of confidentiality. The items were:

1. I believe that a girl who likes to look sexually attractive is unlikely to be a devoted wife when married.
2. I have found (or probably would find) it very difficult to fall in love with a girl with whom I have had intercourse.
3. From my experience, I would say that all women are fundamentally "sexual creatures" and that when they appear to be decent and kind it is only a façade.
4. My feeling that there is something basically animal-like and crude associated with sex interferes (or would interfere) with my expressing myself sexually toward someone I like very much.
5. I find myself losing my respect (or I would lose my respect) for a girl who lets me neck or pet with her on our first date.
6. From my experience I would say that in general very few sexy girls have sweet, decent dispositions.
7. I find that it is easy (or would find it easy) to get sexually aroused with a girl who has had sexual experience, but find it (or would find it) difficult to express my sexual feelings to a girl I love and respect.
8. I feel that basically there are two types of girls: the sexy type and the nice, respectable type.

[4] A copy of the complete scale has been deposited with the American Documentation Institute.

## RESULTS

### *Perception of Sexuality and Affection*

To test the hypothesis that high authoritarians would perceive the star as more sexual than kindly, a mixed Type III analysis of variance (Lindquist, 1953) was employed. The two dependent variables were the scores for each *S* on the sexual and kindness subscales of the movie rating scale. Analysis indicated that regardless of group differences, *S*s as a whole did not perceive the star as being either significantly more sexual ($M$ = 43.47) or more kindly ($M$ = 40.28). Order of presentation of the movies also was found not to influence *S*s' movie ratings. Significant differences at the .025 level, however, were found in *S*s' sexual and kindness ratings when the effects of authoritarianism were evaluated ($F$ = 5.34; $df$ = 1). As can be seen in Table 1, high authoritarians rated the star as significantly more sexual than kindly, while the ratings of low author-

TABLE 1

COMPARISON OF SEXUAL AND KINDNESS SCORES FOR THE HIGH AND LOW AUTHORITARIAN GROUPS

| Authoritarian | *N* | Sexual Scores | | Kindness Score | | *t* |
|---|---|---|---|---|---|---|
| | | *M* | *SD* | *M* | *SD* | |
| High | 32 | 45.72 | 6.39 | 38.69 | 9.45 | 3.41* |
| Low | 32 | 41.22 | 6.53 | 41.88 | 6.29 | .41 |

* Significant at beyond the .001 level for 62 *df*.

itarians were practically identical on the two variables.[5] Further, high authoritarians perceived the star as significantly more sexual than did low authoritarians ($t = 2.73$; $df = 32$; $p = .01$). Inspection of Table 1 also reveals that low authoritarians tended to rate the star as more kindly than did high authoritarians, but this difference fell short of significance ($t = 1.56$; $df = 32$; $p = >.10$).

*Dichotomization of Sexuality and Affection*

Three measures were used to assess the splitting of sexuality and affection. Two of these, ratings of sexuality and kindness, and the organizational aspects of the written impressions, were based on the movies, while the third pertained to *S*s' self-reports of their heterosexual behavior and attitudes.

It was found that the groups high and low in authoritarianism did not differ significantly on the movie measure of splitting, expressed as the absolute difference between each *S*'s sexual and kindness ratings.

The written impressions of the star were classified into organizational type according to Gollin's (1951) criteria. Interjudge agreement based on each of two judges independently classifying 52 impressions was 78.8%. Results presented in Table 2 suggest that authoritarians tended to organize their impressions in a nonintegrated, that is aggregated or unitary manner, relative to the low authoritarians, the majority of whom formed their impressions in a more integrated style. However, the significance of this finding as determined by a 2 × 2 chi square analysis between authoritarianism and integration of impressions (integrated vs. nonintegrated categories) was > .10 ($\chi^2 = 1.82$; $df = 1$).

When the written impressions were analyzed for affective tone (see Table 3), there was again a tendency, though nonsignificant, for high authoritarians to characterize the star more unfavorably than favorably, relative to the low authoritarians. Interjudge agreement, based on 52 randomly selected protocols, independently classified by two scorers, was 80.8%.

The findings for the effects of authoritarianism and sequence of presentation on degree of splitting of sexuality and affection in *S*s' reported heterosexual behavior and attitudes are presented in Tables 4 and 5. As can be seen, both authoritarianism and sequence of presentation were significantly related to degree of dichotomization in *S*s' perceived behavior.

TABLE 2
DISTRIBUTION OF IMPRESSION CATEGORIES FOR HIGH AND LOW AUTHORITARIANS

| Authoritarianism | Organization of Impressions | | |
|---|---|---|---|
| | Integrated | Nonintegrated | |
| | | Aggregated | Unitary |
| High | 7 | 15 | 10 |
| Low | 13 | 11 | 8 |
| *N* | 20 | 26 | 18 |

TABLE 3
CHI SQUARE ANALYSIS BETWEEN AUTHORITARIANISM AND FAVORABLENESS OF IMPRESSIONS

| Authoritarianism | Favorableness of Impressions | | |
|---|---|---|---|
| | Favorable | Unfavorable | Both Favorable and Unfavorable |
| High | 8 | 13 | 11 |
| Low | 14 | 7 | 11 |
| *N* | 22 | 20 | 22 |

Note.—$\chi^2 = 3.44$; $df = 2$; $p = >.10$.

TABLE 4
EFFECTS OF AUTHORITARIANISM AND SEQUENCE PRESENTATION ON THE DEGREE OF "SPLITTING" OF SEXUALITY AND AFFECTION IN REPORTED HETEROSEXUAL BEHAVIOR

| Source | *df* | *SS* | *MS* | *F* |
|---|---|---|---|---|
| Authoritarianism (F) | 1 | 129.40 | 129.40 | 12.72** |
| Sequence Presentation (S) | 1 | 40.65 | 40.65 | 3.997* |
| F × S | 1 | 13.13 | 13.13 | 1.29 |
| Within | 60 | 610.06 | 10.17 | |
| Total | 63 | 793.24 | | |

* $p = .05$ for $F = 4.000$.
** $p < .001$ level.

[5] An *F* test was performed between the variances of the high authoritarians' sexual and kindness scores and heterogeneity of variance was indicated ($F = 2.19$; $df = 31, 31$; $p = .04$). One should regard the obtained *t* value as somewhat inflated. However, since the significance of *t* was beyond the .001 level, it seems safe to conclude that a genuine difference does exist between the means of the sexual and kindness scores.

TABLE 5
HETEROSEXUAL QUESTIONNAIRE SCORES IN RELATION TO AUTHORITARIANISM AND SEQUENCE OF PRESENTATION

| Authoritarianism | Heterosexual Questionnaire Score | | |
|---|---|---|---|
| | *N* | *M* | *SD* |
| High | 32 | 18.03 | 3.47 |
| Low | 32 | 15.19 | 2.94 |
| Sequence of Presentation | | | |
| SK | 32 | 17.41 | 3.19 |
| KS | 32 | 15.81 | 3.64 |

High authoritarians showed a greater tendency to separate sexual impulses and feelings of affection in their reported heterosexual relationships, relative to low authoritarians. *S*s presented with the SK movie sequence also reported that the separated sexuality and affection to a greater extent than *S*s shown the KS sequence.

An insignificant correlation of .046 was obtained between the movie (absolute difference measure) and questionnaire measures of splitting.

## DISCUSSION

The hypothesis that high authoritarians would overemphasize the sexual attributes of the star was confirmed. This result is consistent with the underlying theory of the authoritarian personality. Basic to the evolution of the personality structure of the authoritarian is the unacceptability of id impulses. Sexual strivings remain for the most part suppressed and isolated. However, since they are not integrated with the rest of personality, these impulses find projected and other displaced outlets, as in a heightened concern and moral indignation of sexual expression in others. One can then view the high authoritarian's essentially sexual perception of the star in terms of his projecting unacceptable sensual impulses on to her. However, another factor may be involved. Even when sexual impulses are acted upon by the authoritarian, it is rarely for the purpose, primarily, of genuine sexual gratification. For the male authoritarian, sexual expression tends not to be a goal in itself, but a means by which status is acquired and his sense of masculinity enhanced. His agreement with items suggestive of sexuality may, in part, be attributed to his desire to demonstrate a heightened sexual interest in women as proof to himself as well as others that he is fulfilling his idealized conception of the masculine role.

The hypothesis that high authoritarians would show a splitting of sexuality and affection was partially substantiated. Analysis of the heterosexual questionnaire scale supported the view that high authoritarians possessed dichotomized sex attitudes to a greater degree than low authoritarians. They should therefore find it more difficult, relative to low authoritarians, to react both sexually and affectionately toward the same woman. Instead, authoritarians show a tendency to react sexually toward women they do not respect or admire and inhibit sexual expression when conscious feelings of respect and affection are evoked. Findings on the organization of impressions, while not statistically significant, were in the predicted direction. The one inconsistency was that in their ratings (as determined by the absolute difference between each *S*'s sexual and kindness scores), high authoritarians were not found to separate sexuality and affection to a greater extent than low authoritarians. To obtain a high "splitting index" requires that *S* have a high score on one scale and a low score on the other. The rationale of this measure lies in the assumption that if sexuality and affection could not be integrated, *S* would have to acknowledge only one attribute, the other characteristic being incompatible and hence denied. However, the converse of this, whereby *S* acknowledges and responds to both characteristics, does not necessarily imply integration. Sexuality and affection can both be recognized, yet remain phenomenologically as isolated attributes. The absolute difference score hence does not indicate whether cognitive fusion actually occurred. What was found was that most *S*s did perceive both sexuality and affection in the star. While the high authoritarians did not show splitting, in the terms just outlined, to a greater degree than the lows, we cannot conclude that the same degree of integration was present for both groups. It seems that a more valid index of the integration of sexuality and affection is provided by the integrated category of the written impressions,

which goes beyond the perceptually given and requires merging at a cognitive level.

Although analysis of the written impressions tended to support the hypothesis that high authoritarians would form more nonintegrated impressions, relative to the lows, this finding fell short of statistical significance. One possible explanation is that the star's behavior may not have been sufficiently sexual or kindly to represent extremes or stereotypes. It should be remembered that in the sexual scene, behavior suggestive of promiscuity or immorality was avoided. Again, in the kindness scene, it may have been possible for some to infer a sexualized quality in the star's behavior toward the adult male, even though affection was manifestly revealed. Further, it was impossible to portray only sexuality and kindness, which, while representing the major behavioral characteristics, were confounded to some extent with other attributes. Slovenliness or boredom was hinted at in the sexual scene and intelligence and cultivation in the kindness scene. In addition, the content of several of the items of the sexual scale implied a "looseness" beyond mere sexual attractiveness. *Ss* thus may have been reacting to other qualities apart from the star's kindly and sexual features, even though inspection of the written impressions revealed that the great majority focused primarily on the two major attributes.

Viewed as a whole, the present findings substantiate one of the underlying hypotheses of *The Authoritarian Personality* in that high authoritarians were found to over-react to sexuality in women. While present results also support Frenkel-Brunswik's conclusions regarding the male authoritarian's tendency to separate sexuality and affection in actual behavior, the phenomenon of nonintegration was only equivocally demonstrated in the experimental task. The results are also significant in that they demonstrate the relevance of authoritarianism to men's social perception of women.

## SUMMARY

The present study investigated the effects of authoritarianism on men's reactions to sexuality and kindness in women. Sixty-four college males served as subjects, half the group consisted of high authoritarians and the other half comprised the low authoritarian group. The subjects were shown a silent movie consisting of two scenes, one sexual and one kindly, in which a young, attractive woman appeared. For half the group the sexual scene was shown first, and for the other half the kindly scene was first presented.

After seeing the movie, subjects wrote their impressions of the young woman's (star's) personality, after which they evaluated the star on a specially devised rating scale. The subjects then answered a questionnaire relating to their heterosexual behavior and attitudes.

The high authoritarian group perceived the star as significantly more sexual than kindly and also reported "splitting" of sexuality and affection in heterosexual behavior to a significantly greater degree than was reported by low authoritarians. The highs, in comparison to the lows, tended to form unfavorable impressions of the star, in which sexuality and kindness were not integrated. However, this difference was not statistically significant.

## REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.

GOLLIN, E. S. Forming impressions of personality. Unpublished doctoral dissertation, Clark University, 1951.

GOLLIN, E. S. Forming impressions of personality. *J. Pers.*, 1954, **23**, 65–67.

GOLLIN, E. S. Organizational characteristics of social judgements. A developmental study. *J. Pers.*, 1958, **26**, 139–153.

GOLLIN, E. S., & ROSENBERG, S. Concept formation and impressions of personality. *J. abnorm. soc. Psychol.*, 1956, **52**, 39–42.

JONES, E. E. Authoritarianism as a determinant of first impression formation. *J. Pers.*, 1954, **23**, 107–127.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.

SCODEL, A., & FREEDMAN, M. Additional observations on the social perceptions of authoritarians and nonauthoritarians. *J. abnorm. soc. Psychol.*, 1956, **52**, 92–95.

SCODEL, A., & MUSSEN, P. Social perceptions of authoritarians and nonauthoritarians. *J. abnorm. soc. Psychol.*, 1953, **48**, 181–184.

THIBAUT, J. W., & RIECKEN, H. W. Authoritarianism status and the communication of aggression. *Hum. Relat.*, 1955, **8**, 95–120.

(Received June 29, 1959)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 335-340

# GROUP INTERACTION AND PERCEIVED SIMILARITY OF MEMBERS[1]

EZRA STOTLAND, NICKOLAS B. COTTRELL
*University of Washington*
AND GORDON LAING
*University of Michigan*

THE problem of the present study is the relationship between the amount of interaction in a group and the members' tendency to perceive other members as being similar to themselves. The general hypothesis is that members of a group in which there is a high degree of interaction perceive more similarity between themselves and other members than members of a group in which there is a low degree of interaction. The validity of this hypothesis is assumed to be limited to those cases in which the interaction conveys no information about the characteristics of the group members.

The theoretical source of this hypothesis is in the work of Michotte (1946). His studies of the perception of causal relationships between physical objects suggest that when one object is perceived to cause a change in a second object, the change in the second is perceived to belong to the first. In other words, a new unit or new entity is perceived, consisting of the first, causing object and of the change in the second object. This proposition has been extended theoretically to the area of social psychology by Heider (1958) who postulates that a person and an event he has been perceived to cause tend to be perceived as a single unit.

These ideas can be extended theoretically to the area of the perception of the interaction between two (or more) persons. By definition when two people interact, each act by one of them is followed by or causes some event (e.g., change in behavior) in the other. The causing person and the event in the other person should then constitute a perceptual unit. A series of such units should be formed as each person acts so as to cause an event (or reaction) in the other person. It is not unreasonable to assume that these series of units cumulate or synthesize to form a composite single unit, since each unit in the series is constituted by the same two (or more) persons. Such a synthesis would be expected from the laws of cognitive organization, which hold that, in general, the perceiver attempts to achieve the simplest possible organization of cognition. The composite units formed in interaction can vary in the number of interactions upon which they are based. Let us term this difference as the degree of strength of the composite unit, the larger the number of individual units, the greater the strength. The idea of strength is analogous to the concept of strong and weak gestalts.

One additional assumption is needed to derive the hypothesis of the present study. The phenomenon of assimilation found in the study of the perception of physical objects is the tendency to perceive all parts of the same perceptual unit as having the same quality, e.g., degree of brightness. It is reasonable to expect that this assimilative tendency toward uniformity of parts of a unit would be greater in strong units than in weaker units. As mentioned above, the number of interactions between two or more persons defines the strength of the resultant unit that they generate. Therefore, it would be expected that when there are many interactions between or among them, there should be a greater tendency to perceive them as uniform.

Perception of uniformity can arise in a given member of a group of two or more persons either by his perceiving that the other members have characteristics similar to ones he already believes he has, or by perceiving himself to have characteristics that he previously assumed the other persons to have. The present study is limited to the attribution of own characteristics to others. Our hypothesis is that in groups in which there are a high number of interactions among members, each member should attribute more of his own characteristics to the other members than in groups in which

[1] The work on this study was financed by a grant (M-2423) from the National Institutes of Health of the United States Public Health Service.

there are a low number of interactions. The underlying process is assumed to occur independently of the perception of similarity of members resulting from their acceptance of the group norm or standard, and independently of any perception of other members that may result from the interchange among the members of information regarding themselves.

## METHOD

*Subjects.* *Ss* were 90 students, 58 men and 32 women, required to participate as an assignment in an introductory psychology course at the University of Washington. In assigning *Ss* to six-person experimental groups, an effort was made to assign *Ss* from different living groups, discussion sections, majors, etc. to minimize the possibility of previous mutual acquaintance. Each group consisted of like-sexed persons.

*Procedure.* *Ss* arrived in the experimental room in groups of six. They were asked to sit around a small round table which had a Lazy Susan in the center with six compartments. Each *S* could see only the faces of the other *Ss*, since there were wooden separations between them. In the compartment confronting each *S* was a pack of six slips of paper with 10 syllables on each one; a pack of six slips with five lines on each numbered from 1 to 5; a sheet with a title at the top and with two columns of 30 letters in all, in alphabetical order; a red cardboard; a pencil; and a pad of scratch paper.

*Ss* were told that they were to list names from different geographical areas of the world, each person collecting names from a different area. Each *S* found the name of his area at the head of the sheet with the two columns of letters and also was given a list of all the areas (North America, Pacific, Latin America, Asian Mainland, European countries, European cities). There were world maps delineating these areas on the walls of the room. In addition, each of the compartments in the Lazy Susan was labeled for a given area.

*Ss* were told that they were to form geographical names by combining syllables in pairs collected from sets of 10 syllables they found on the slips before them. Thus, each *S* produced five names by trying various combinations of the 10 syllables. He listed these names first on one of the sheets with five lines numbered 1 to 5 (although many *Ss* did not obey the instructions to list the names first on the sheets). *S* was to write any geographical name from his own area on the sheet with the columns of letters, after the initial letter of the name he had formed. The objective was to find one name for each of the letters on his sheet. If *S* formed a name from an area other than his own, he was to write it on a slip of scratch paper, place it in the appropriate compartment in the Lazy Susan, and then turn the Susan until the right *S* received it. There was no trouble resulting from several *Ss* desiring to use the Lazy Susan at the same time. Each *S* had six such sets of 10 syllables. Thus, any *S* could complete his collection of 30 names from names he had formed himself, as well as from names he received from others. If he found he was short a name after he had completed work on his syllables, he placed the red cardboard in his compartment so that other *Ss* would notice that an error in sending a name had occurred, and correct it.

Few *Ss* were actually able to complete their lists of 30 names, apparently because of the concoction of nonexistent names by some *Ss*, lack of intellectual capacity, and failure to reject erroneous names received. Groups were allowed to continue until their functioning broke down as a result of the concoction of names. No group was allowed to run longer than 26 minutes on the task; the shortest time was 21 minutes. After data had been collected from 31 *Ss*, a record was kept of the number of names completed by each remaining *S*, regardless of whether they were sent to other *Ss* or returned. As will be shown below, the number of names completed had no effect on the dependent variable.

It is obvious that the task just described could be arranged so that each *S* would work mostly on names from his own area or mostly on ones from other *Ss*' areas. In the *high interaction* condition, 25 of the 30 names each *S* formed from the syllables were from an area other than his own. He would have to send five names to each of the five other *Ss*. He would, in turn, receive the 25 names he needed from the other *Ss*. In the *low interaction* condition, 25 of the 30 names were from each *S*'s own area. He would send the other five to the other five *Ss*, one to each, receiving one from each in return. The groups were assigned alternately to one or the other condition.

The task thus permitted variation in the degree of interaction among the *Ss*, with about the same amount of work by each *S*. Each *S*'s contribution to the interaction could be equated, and the interaction conveyed no information about the *Ss*' personalities.

When the task was completed, *Ss* were given a list of 30 items from a variety of personality scales (e.g., general anxiety, hostility, personal history, test anxiety),[2] selected so as to refer to information about which students would not in general be likely to communicate with each other, and thus minimizing the possibility that *Ss* would use their knowledge of typical student feelings and behavior in predicting other *S*'s responses. The *Ss* first indicated whether each of these statements was true with regard to themselves. Then they were asked to respond to an identical list on a separate sheet, guessing how the majority of the other *Ss* in the group would respond to each item. The dependent variable was the number of agreements between own responses and predictions of others' responses. Finally, *Ss* filled out a post-experimental questionnaire with 7-point scalar items.

## RESULTS

Before the hypothesis of the study can be tested, it is necessary to show that the high and low interaction conditions differed in no way (except in the degree of interaction) that could explain differences in the levels of the

[2] The authors wish to thank Irwin Sarason for his assistance in obtaining these items.

dependent variable. In Table 1 are presented the mean differences between the conditions in items of the post-experimental questionnaire. It can be seen there that the two conditions did not differ in degree of attraction to the group or to its members, interest in the various experimental tasks, perception of group "performance," or perception of integration into a "close-knit" group. However, there tended to be a higher degree of prior acquaintance among the *Ss* in the low interaction condition than in the high interaction condition, a difference that must be taken into account in subsequent analyses.

The hypothesis of the study was that in the high interaction condition, the *Ss* would predict more similarity between themselves and the other *Ss* than in the low interaction condition. There was no significant difference between the two conditions in this respect. For the high interaction condition the mean was 22.74, for the low, 22.53.

However, a clearer picture of the results is obtained by taking into consideration two other factors which can influence perception of own similarity to others, i.e., attraction to others and prior acquaintance with them.

TABLE 1

MEAN RESPONSES OF SUBJECTS TO ITEMS IN THE POST-EXPERIMENTAL QUESTIONNAIRE

| Item | Experimental Condition | |
|---|---|---|
| | High Interaction ($N = 45$) | Low Interaction ($N = 44$) |
| 1. Degree of prior acquaintance with other subjects[a] | 7.27* | 8.57* |
| 2. Desire to get to know other subjects[a] | 19.91 | 19.38 |
| 3. Interest in geography game | 5.56 | 5.40 |
| 4. Interest in filling in personality questionnaire regarding self | 4.29 | 4.18 |
| 5. Interest in predicting other subject's responses to personality questionnaire | 4.23 | 4.00 |
| 6. Evaluation of group performance on geography game | 3.75 | 3.56 |
| 7. Degree of integration of group into a close-knit group | 3.44 | 3.18 |
| 8. Desire to leave present group to join another working on the same task | 3.36 | 3.36 |

[a] Computed by having the *S* rate his degree of prior acquaintance with (Item 1) or his desire to know (Item 2) each of the other *Ss* and summing up these ratings.

* Significantly different by a two-tailed *t* test, $t = 2.043$, $p < .05$. A high number indicates greater prior acquaintance. No other means were significantly different between conditions.

TABLE 2

MEAN NUMBER OF PREDICTIONS OF OTHER SUBJECTS' AGREEMENTS WITH SELF FOR SUBJECTS OF NO PRIOR ACQUAINTANCE ($N = 6$ in each cell)

| Desire to Know | Experimental Condition | | Total |
|---|---|---|---|
| | High Interaction | Low Interaction | |
| High | 25.00 | 22.50 | 23.75 |
| Low | 22.17 | 19.57 | 20.87 |
| Total | 23.58 | 21.04 | |

Table 2 presents the mean number of predictions of similarity for *Ss* in each experimental condition with no prior acquaintance with other *Ss*, controlling on the desire to know the *Ss*. The analysis of variance, shown in Table 3, indicates that experimental conditions, as well as the desire to know other *Ss*, have highly significant effects on the number of predictions of similarity of other *Ss* to self. In Table 2, it can be seen that the major hypothesis of the study is confirmed, since the mean for the high interaction condition is significantly greater than for the low. A similar comparison of the conditions for *Ss* who had prior acquaintance produced no significant difference. The hypothesis holds only if the *Ss* have no other information about the other *Ss* because of prior acquaintance. Among these *Ss*, the desire to know other *Ss* is associated with prediction of similarity, as would be expected from other research.

However, it is possible for the findings to be explained by a process quite different from the one outlined above. The first step in the process would be for the *Ss* in the high interaction condition to develop more liking for each other than those in the low interaction condition. The greater liking would then lead the *Ss* in the high interaction condition to attribute more desirable attributes to the others. Since in general persons tend to attribute desirable traits to themselves, the outcome of this process would be greater similarity in the high interaction condition between the *Ss*' self-descriptions and his descriptions of the others than in the low interaction condition. This is precisely the result obtained.

This explanation in terms of the influence of

TABLE 3

ANALYSIS OF VARIANCE OF PREDICITIONS OF OTHER SUBJECTS' AGREEMENTS WITH SELF FOR SUBJECTS OF NO PRIOR ACQUAINTANCE

| Source | *SS* | *df* | *MS* | *F* |
|---|---|---|---|---|
| Experimental Condition | 42.666 | 1 | 42.66 | 11.91** |
| Desire to Know | 54.000 | 1 | 54.000 | 15.07** |
| Experimental Condition × Desire to Know | .166 | 1 | .166 | |
| Within cells | 71.668 | 20 | 3.583 | |
| Total | 168.500 | 23 | | |

** $p < .01$.

interaction on attraction would, however, be difficult to reconcile with the results of Tables 2 and 3, since these tables show that there is more prediction of similarity in the high interaction condition than in the low, independent of the degree of attraction to other members. If the attraction explanation were valid, it would be expected that there would be no difference between the conditions independent of the degree of attraction.

Another way of resolving the issue between the "attraction" explanation and the one described earlier in the paper is to determine whether the *S*s in the high interaction condition tended more than those in the low interaction condition to predict that the other *S*s in their groups were similar to themselves with regard to socially desirable traits. If the difference between conditions occurred only with regard to the socially desirable traits, the attraction explanation would be supported. If the difference between conditions occurred equally for both desirable and undesirable traits, then support would be given to the explanation in terms of greater unit formation with greater interaction.

The following procedure was used to resolve this issue: 75 undergraduate *S*s of the same general background as those in the present study sorted the statements used in the personality questionnaire into six equal piles in the order of desirability of the statement's being true about other people. These rankings were pooled, and the total sample of statements was then split at the median rank on social desirability.

This split of the items on social desirability was then used to analyze the results of the original study. A socially desirable *response* was defined as an answer of "true" to a socially desirable statement or as an answer of "false" to a socially undesirable statement. For those statements to which an *S* gave a socially desirable response, the number of his predictions of similar responses by other *S*s was calculated. A similar total was calculated for those statements on which an *S* made an undesirable response about himself.

It is obvious that a person who attributes a high number of socially desirable traits to himself will have more opportunities to predict that others have the same socially desirable traits than will a person who attributes a low number to himself. Therefore, a count was made of the number of socially desirable responses each *S* made with regard to himself and the *S*s were split into two groups at the median. The mean numbers of predictions of similar socially desirable and undesirable responses were then calculated separately for the high and low social desirability *S*s. These means were calculated for the high and low interaction conditions, employing only those *S*s who had no prior acquaintance. The results are presented in Table 4.

This table shows no difference between the experimental conditions in the number of own socially desirable responses attributed to others, regardless of the number of own socially desirable responses. This lack of difference does not support the interpretation in terms of difference in attraction. On the other hand, a difference between the conditions significant at the .01 level is obtained for the prediction of socially undesirable responses made by *S*s low in the social desirability of their self-descriptions. In the high interaction condition, these *S*s had a greater tendency than in the low interaction condition to perceive the other *S*s as having the same undesirable traits as themselves. This difference is therefore completely inconsistent with the interpretation in terms of the effect of interaction on liking. Thus, both these data and the previously cited independence of the effects of attraction and of the experimental conditions argue strongly against the attraction interpretation. However, these results do not fit the "unity" explanation very well either.

The results just described appear to be inconsistent with Homans' (1950) hypothesis

that greater interaction leads to greater liking. However, another test of the Homans hypothesis can be made by examining the difference between the conditions in attraction to the other members separately for *S*s high and low in the social desirability of their self-descriptions. In Table 5 are presented the appropriate mean post-experimental ratings of desire to know other *S*s, for the *S*s of no prior acquaintance. None of the differences between conditions is significant. Nonetheless, the general trend is consistent with the Homans hypothesis, especially for *S*s low in the social desirability of self-descriptions, who are more attracted to the other *S*s in the high interaction condition than in the low. However, these *S*s also predicted that the others were more like themselves in undesirable traits in the high interaction condition. It appears then that for these *S*s the similarity of traits was more significant for the degree of attraction to others than was the desirability of others' traits. Therefore, Homans' hypothesis may be valid with regard to attraction, but not with regard to favorability of perception.

TABLE 4

MEAN NUMBER OF PREDICTIONS OF SOCIALLY DESIRABLE AND UNDESIRABLE RESPONSES BY OTHER SUBJECTS SIMILAR TO RESPONSES WITH REGARD TO SELF FOR SUBJECTS OF NO PRIOR ACQUAINTANCE (*N*s in parentheses)

| | Experimental Condition | |
|---|---|---|
| | High Interaction | Low Interaction |
| **I. *S*s high in attribution of socially desirable traits to self** | | |
| Desirable perception of self | 16.8 (5) | 16.2 (6) |
| Undesirable perception of self | 6.00 | 6.50 |
| Total | 22.80 | 22.70 |
| **II. *S*s low in attribution of socially desirable traits to self** | | |
| Desirable perception of self | 13.1 (8) | 11.7 (6) |
| Undesirable perception of self | 10.0 (A) | 7.50 (B) |
| Total | 23.1 (C) | 19.2 (D) |

Note.—Cells significantly different by two-tailed *t* test:
A vs. B ($t = 3.157$, $p < .01$).
C vs. D ($t = 3.900$, $p < .01$).

TABLE 5

MEAN RATINGS OF DESIRE TO KNOW OTHER SUBJECTS FOR SUBJECTS OF NO PRIOR ACQUAINTANCE (*N*s in parentheses)

| *S*s' Attribution of desirable traits to self | Experimental Condition | |
|---|---|---|
| | High Interaction | Low Interaction |
| High | 21.1 (7) | 18.4 (8) |
| Low | 21.3 (6) | 15.8 (9) |
| Total | 20.1 | 17.5 |

Note.—No cells significantly different.

It was found that more names were composed by *S*s in the low interaction condition than in the high, the mean for the low interaction condition being 26.21 names, which was significantly different at the .001 level ($t = 3.04$) from the mean of 21.73 for the high interaction condition. (This mean is based only on the 59 *S*s for whom these records were kept.) The reason for this discrepancy probably is that the world-wide frame of reference necessary in the high interaction condition troubled some *S*s. The question may be raised whether the number of words composed by an *S* affected his prediction of other *S*s' agreement with him in the two experimental conditions. However, when the conditions were compared separately for *S*s above and below the median in number of names formed, these two groups of *S*s were found not to differ in their reaction to the condition differences.

## DISCUSSION

The results showed that the experimental hypothesis, that the degree of perception of similarity of others to self is a function of the degree of interaction between self and others, was supported only among those *S*s who had no prior acquaintance with other *S*s. It seems that when the *S*s have some knowledge of the other *S*s, they use this information in their judgments of the other *S*s. The more subtle process of projecting own traits onto others, as a result of interacting with them, can occur only if such information is unavailable. In addition, the results show that the interaction leads to more perception of similarity mostly by *S*s who attribute relatively more socially

undesirable traits to themselves, and then by perceiving the others to be similar to self on these undesirable traits. It may be that these *S*s are more reactive to differences in social situations as has been found in other studies (e.g., Stotland, Thorley, Thomas, Cohen, & Zander, 1957). However, it is difficult to understand why they perceived the other persons in the high interaction condition as being more similar to them in undesirable ways. These results raise questions about the adequacy of explaining the findings in terms of more perception of unity of the group arising from greater interaction, as was postulated in the original theory.

Other factors as well as interaction can, of course, increase the perception of similarity between self and other persons. In the present study, perception of such similarity was associated with a high attraction to other persons, a finding consistent with other research. Another factor which could lead to perception of similarity would be the development of a group norm. However, the development of a norm is contingent upon the exchange of information of some sort among the group members, which was of course impossible in the present experiment.


## SUMMARY

It was hypothesized that persons with a high degree of interaction among themselves would perceive more similarity among themselves than persons with a low degree of interaction. It was predicted that this hypothesis would be supported when the interaction did not convey any information relevant to the degree of similarity among the subjects. This hypothesis was tested in a laboratory experiment. The findings were: (*a*) that the hypothesis was supported among those subjects with no acquaintance prior to the experiment; (*b*) that there was an association between perception of similarity to other subjects and desire to get to know them.


## REFERENCES


HEIDER, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

HOMANS, G. C. *The human group.* New York: Harcourt, Brace, 1950.

MICHOTTE, A. E. *La perception de la causalité.* Paris: Vocin, 1946.

STOTLAND, E., THORLEY, S., THOMAS, E. J., COHEN, A. R., & ZANDER, A. The effects of group expectations and self-esteem upon self-evaluation. *J. abnorm. soc. Psychol.*, 1957, **54**, 55–63.


(Received July 2, 1959)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 341-347

# VERBAL RATE, EYEBLINK, AND CONTENT IN STRUCTURED PSYCHIATRIC INTERVIEWS[1]

FREDERICK H. KANFER

*Purdue University*

RECENTLY, several attempts have been made to relate the momentary changes in subject's (*S*'s) anxiety level during the interview to structural speech characteristics (Dibner, 1956; Mahl, 1956), to GSR (Dibner, 1958; Dittes, 1957), and to respiratory activity (Goldman-Eisler, 1955, 1956). Panek and Martin (1959) also found a high degree of coincidence of speech disturbances and GSR during interviews. The primary purpose of these studies has been to determine the validity of a construct of momentary anxiety and its indices in order to understand interview behavior more completely. In the final analysis, demonstration of specific relationships of these indices to the momentary *content* of the interview would also lead to refinement of interviewing techniques and more adequate use of the interview as a method in personality research.

In a series of earlier studies, verbal rate was related to some experimentally controlled modification of *S*'s momentary emotional state, either by conditioning (Kanfer, 1958) or by selecting topics to which *S* was differentially sensitive (Kanfer, 1959). The earlier findings suggested that verbal rate increases with *S*'s present anxiety level and that a relatively stable relationship exists between verbal rate and content of speech. The present study attempted to extend the generality of the earlier findings by introducing a series of conceptually related external criteria which reflect *S*'s current and long-standing emotional state. For the present study a group of *S*s was selected, whose emotional adjustment was wider in range, less stable in its momentary content-related fluctuations, and more sensitive to separate content areas than the college population on which the earlier research was done. Since interview behavior is undoubtedly a function of the importance of the interview outcome in *S*'s current life processes, research with volunteer college students attenuates the effectiveness of any experimental variable relating to the complex emotional life history of *S*. Therefore, the present study used structured interviews which were imbedded in the routine admission procedures of a psychiatric hospital and were perceived by *S* as highly important in his present life situation.

The present study was designed to investigate several related hypotheses. The *first* concerned the relationship between verbal rate and content. In selecting the content for the interview it was predicted that any topic which is clearly of great emotional impact to all *S*s would result in significantly higher verbal rate than a topic which is not immediately related to *S*'s present emotional problems. It was hypothesized that a direct question about *S*s' reason for coming to the hospital would yield a significantly higher mean verbal rate than questions relating to other aspects of their life.

The *second* hypothesis dealt with the evaluation of *S*'s momentary anxiety by a concurrent nonverbal measure. If high verbal rate is related to increased momentary anxiety, then some nonverbal measure of *S*'s tension level during the discussion of various interview topics would show similar changes.

The human eyeblink has been postulated to reflect many changes in a person's overt and covert activity. Meyer (1953) proposed blink rate as a measure of muscular tension, and Doehring (1957) and Meyer, Bahrick, and Fitts (1953) have shown its relationship to some anxiety measures. Since it can also be recorded without affecting *S*'s interview behavior, this measure was selected for the present study.

The second hypothesis may now be stated more specifically. If the predicted changes in speech rate over topics are related to *S*'s momentary anxiety level, then blink rate is expected to show similar differences with the highest blink rate on the topic which arouses

[1] This study was supported in part by Research Grant M-2027 (C1) from the National Institute of Mental Health, United States Public Health Service. The writer wishes to acknowledge the contributions of Ann Durkee-Riggs who served as *E* and assisted in the execution of the study.

the greatest tension. Although high correlations between blink rate and verbal rate are not expected because these variables differ in their *specific* determinants and the measures differ in their sensitivities, the general pattern of rate changes within topics is expected to be similar.

It has been suggested earlier (Kanfer, 1959) that verbal rate may relate to *immediate* changes in *S*'s emotional state even though an anxiety measure which reflects a lasting personality attribute does not allow differential predictions of verbal rate for individual *S*s on unselected topics. Nevertheless, speech and eyeblink patterns may be associated with other personality attributes measured by psychological tests, psychiatric ratings, and diagnostic classification. In order to test this suggestion and to provide grounds for better conceptualizations in this area, the general relationship between verbal rate, eyeblink, personality test scores, and psychiatrist ratings was also explored.

## METHOD

*Subjects.* Thirty-eight married female psychiatric patients between the ages of 21 and 36 served as *S*s in this study. The mean age was 28.7 years. All patients were assigned to the study immediately after admission to the hospital, except those who were too disturbed to take the test battery, whose admission diagnosis suggested organic factors, or who were started on treatment before they could be interviewed for this study.[2] The following diagnostic categories were represented among the *S*s: acute schizophrenia (12), paranoid schizophrenia (9), depressive reaction (3), dissociative reaction (2), anxiety reaction (3), obsessive-compulsive (2), unstable and inadequate personalities (6), and schizoid personality (1). Two *S*s were later eliminated from the study due to their inability to complete the interview.

*Apparatus and Procedure.* A test battery consisting of the Taylor Manifest Anxiety Scale (MAS), the Cornell Index, and individual MMPI forms was administered as part of the hospital testing routine to groups of three or four *S*s. On the same day or the following day *S*s were seen individually for the structured interview.

The interviews were conducted in a soundproof room, with *S* facing the experimenter (*E*) across a table and sitting next to a one-way vision screen. A microphone was on the desk. A modified desk clock faced *E*. In the observation room, an observer (*O*) monitored a tape recorder and a timing device which fed a 30-second marker into the recorder and simultaneously activated *O*'s chronoscope. This arrangement permitted synchronization of the recorded verbal material and the eyeblink count made by *O*. Onset of time periods was marked by *O* after *E* completed a question in the structured interview. An electric counter was used to record eyeblinks for 30-second intervals during *S*'s response to each topic.[3]

After *S* was seated in the interviewing room, *E* gave the following instructions:

> I am interested in understanding how you feel about certain things so that we will be better able to help you. I will need as much information and as accurate an account as you can give. Therefore, you can help me if you answer my questions by talking about everything which comes to mind when I raise a topic. Tell me about your attitudes and feelings and give whatever thought or example comes to mind. It is important that you omit nothing. I will not be able to answer any questions after you begin talking. OK? Do you have any questions before we begin?

Four warm-up questions were given, during which *S* was encouraged to talk continuously. The questions for the warm-up were selected on the basis of data from pilot interviews. On these questions *S*s talked freely and did not introduce discussion of their illness or other topics raised in the experimental interview itself. After it was clear that *S* understood the instructions, the following questions were asked and *S* was allowed to talk for at least 2 minutes. After each question *E* summarized *S*'s remarks and proceeded to the next question. For half of the *S*s, the order of Topics B and C was reversed. The questions were:

> A. Where do you live? Tell me about it. How long have you lived there? What do you think of it? What is it like?
>
> B. Tell me a little bit about your family, how you get along with them, and what they are like.
>
> C. Do you feel that you generally have been pretty successful in handling members of the opposite sex, and that you are fairly attractive to men?
>
> D. Do you think you can tell me a little about the trouble that brought you here?

The interviewing and eyeblink recording procedures were practiced with several *S*s in the pilot study. During the interview *E* refrained from any motor activity which could be interpreted as supporting, agreeing, or disagreeing. *E* made no verbal response of any kind during the 2-minute interval.

A rating scale with detailed instructions was given to each *S*'s psychiatrist before he had access to any test findings on the patient. The psychiatrists had no knowledge of the experimental design and were told only that each patient would be interviewed briefly. They were asked to indicate the degree of conflict which they saw in the patient in her relationship with the family, her relationship with members of the opposite sex, and her attitude toward her present illness. Specific examples were given for the extremes of a seven-step scale. In addition, the psychiatrist was asked to indicate on which of the topics the patient "would be most uncomfortable if she had to discuss it in an interview."

---

[2] The patients were interviewed at Larue Carter Memorial Hospital, Indianapolis. The author wishes to acknowledge the cooperation of Donald F. Moore and the hospital staff in contributing their time and facilities to the execution of the experiment.

[3] The author wishes to acknowledge the assistance of Helen Fowler who served as *O*.

TABLE 1
MEAN VERBAL RATES FOR TOPICS AND PERIODS
(In words 30 seconds)

| Periods | Topics | | | |
|---|---|---|---|---|
| | A | B | C | D |
| 1 | 55.81 | 55.47 | 48.64 | 59.33 |
| 2 | 53.78 | 57.92 | 48.52 | 63.35 |
| 3 | 47.18 | 56.08 | 43.23 | 58.22 |
| 4 | 42.95 | 51.94 | 40.81 | 58.61 |

The test batteries were then scored by hospital clerical personnel under the supervision of *E*.

## RESULTS

Topics B and C were given as the second or third topic in the interview, alternated for half of the *S*s. To determine the effect of order, *t* tests were run on verbal rate and eyeblink comparing the two subgroups on each topic. The tests yielded no significant differences between the subgroups. Therefore, all *S*s were pooled in subsequent analyses.

### *Verbal Rate*

Verbal output was measured in words/30 seconds on all topics and treated by analysis of variance. The mean verbal rates on all topics are presented in Table 1. The results of the statistical analysis are shown in Table 2. Verbal rate differed significantly from topic to topic for all *S*s. The highest verbal rate was obtained on Topic D (Illness) which had been selected a priori for its high anxiety-evoking potential for all *S*s. A Duncan analysis on the groups indicated that all means differed significantly from each other at the .05 level. The findings thus support the hypothesis that verbal rate varies with topics and that the highest rate occurs on a topic which is anxiety-arousing to all *S*s and which represents an immediate and realistic problem in the current interpersonal setting.

Verbal rate decreased significantly over periods *within* each topic (Table 2). The nonsignificant topic × period interaction suggests that this pattern of decline was similar for all topics. Overall verbal rate did not decrease simply as a consequence of increased time spent with the interviewer. In fact, Topic D, the last topic introduced, showed the highest verbal rate. In addition, as has been mentioned above, interchanging Topics B and C for half of the *S*s did not result in rate differences. Thus, the rate decrease seems to occur within each topic, and the rate may increase again at the start of a new topic.

### *Eyeblink Rate*

*O*'s reliability of counting blinks was checked in several pilot interviews against two other judges. The operational definition of a blink was based on a complete eyelid closure. The practice resulted in near-perfect agreement of judges' counts for segments of 2 minutes on the pilot interviews.

The summary of the analysis of variance on eyeblinks is presented in Table 3. The results show that blink did not differ significantly over topics for all *S*s, although the topic means array is in the expected direction. Mean blink rates per 30 seconds for topics were: Topic A, 8.95; Topic B, 8.90; Topic C, 8.72; and Topic D, 9.65. If the triple interaction is used as an error term, although use of this estimate results in a lowered $F$ when the variance component for $T \times P \times S \neq$ zero, then the blink rate changes over topics among *S*s can be tested by $T \times S/T \times P \times S$. The significant $F$ ($p < .01$) indicates that the topics differentially affected *individual* *S*s' blink rates. While some *S*s varied in blink rate over topics, all *S*s did not do so consistently.

Since Topic D had been specifically selected for its anxiety arousing potential, the second hypothesis was further tested by a *t* test for correlated measures between each *S*'s mean blink rate on the sum of Topics A, B, and C vs. Topic D. This analysis yielded a *t* value of 1.66, with 35 *df*, which, for a one-tailed test, falls between $p > .05$ and $p < .10$. The results do not give clear-cut support to the hypothesis

TABLE 2
ANALYSIS OF VARIANCE SUMMARY ON VERBAL RATE FOR 30-SECOND PERIODS

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| Subjects (*S*) | 35 | 3956.15 | |
| Topics (T) | 3 | 5739.90 | 22.04** |
| Periods (P) | 3 | 1781.30 | 5.77** |
| T × P | 9 | 203.19 | 0.82 |
| T × *S* | 105 | 260.47 | |
| P × *S* | 105 | 308.76 | |
| T × P × *S* | 315 | 247.61 | |
| Total | 575 | | |

** $p < .01$

that blink rate increases for all *S*s when the interview centered about their illness. However, the data suggest that *S*s show differential blink rates on topics, with the specific pattern determined by individual differences. Furthermore, Topic D tends to yield higher blink rates than the *average* rate on the other topics.

The mean blink rate decreased significantly over periods within each topic. Blink rate increased again whenever a new topic was introduced. The mean blink rates, summed over all topics, were: Period 1, 9.82; Period 2, 9.15; Period 3, 8.75; and Period 4, 8.65.

It was hypothesized that both verbal rate and blink rate are related to *S*'s tension level during the interview. Personality theory suggests that tension in interpersonal situations varies with diagnostic groups. Specifically, the more extensive the patient's disorganization and the greater his tendency to use withdrawal from the environment, the less is the expected tension in an interview. The present sample could be broken down by diagnosis into three groups: acute schizophrenics, paranoid schizophrenics, and others, composed mainly of neurotics and character disorders. The neurotic group would be expected to yield the highest total word rate and blink rate, and the acute schizophrenics to yield the lowest. This hypothesis was tested by an analysis of variance over the three diagnostic groups for total verbal rate and blink rate. The means are shown in Table 4.

For total verbal rate, a nonsignificant *F* ratio of 1.78 (2,33 *df*) was obtained, indicating no evidence that the diagnostic groups differed from each other in verbal rate. The eyeblink data were tested for homogeneity of variance in the groups by the $F_{max}$ test (Walker & Lev, 1953). The group composed of neurotics and character disorders showed the greatest variability and the acute schizophrenics had the lowest variance. The resultant $F_{max}$ of 2.5 was not significant, and an analysis of variance was run. The *F* ratio failed to show significant differences between groups.

TABLE 3
ANALYSIS OF VARIANCE SUMMARY ON EYEBLINK RATE FOR 30-SECOND PERIODS

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| Subjects (*S*) | 35 | 419.68 | |
| Topics (T) | 3 | 24.29 | 0.88 |
| Periods (P) | 3 | 47.39 | 3.77* |
| T × P | 9 | 12.73 | 1.19 |
| T × *S* | 105 | 27.70 | 2.58** |
| P × *S* | 105 | 12.58 | 1.17 |
| T × P × *S* | 315 | 10.74 | |
| Total | 575 | | |

* $p < .05$.
** $p < .01$.

TABLE 4
TOTAL VERBAL AND BLINK RATE MEANS AND STANDARD DEVIATIONS BY DIAGNOSTIC CATEGORIES

| | Acute Schizophrenics ($N = 11$) | | Paranoid Schizophrenics ($N = 9$) | | Neurotics ($N = 16$) | |
|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Verbal | 820.55 | 219.96 | 841.86 | 233.60 | 853.81 | 272.16 |
| Eyeblink | 119.64 | 59.19 | 152.33 | 70.25 | 158.13 | 96.12 |

*Relationship between Psychiatric Ratings, Test Scores, Verbal Rate, and Eyeblink*

Correlations were computed between raw scores on eight of the MMPI scales, the MAS, the Cornell Index, and the psychiatric ratings, and separate verbal rate and eyeblink measures for periods, topics, and verbal and blink rate totals over the interview. Only those correlations which relate the dependent variables, i.e., verbal rate and blink rate, to the various criteria are discussed here. The values of the correlations are presented in Table 5.

Examination of the correlation matrix revealed that all MMPI measures, except the *Hy* and *Ma* scale, showed some relationship to verbal rate during the initial 30 seconds of talk on each topic. The direction of these relationships suggests that *S*s who give a large number of deviant responses on the MMPI scales show a lower initial verbal rate in the interview.

A second cluster of correlations can be seen in Table 5 between psychiatric ratings and blink rate. *S*s who were judged to show more severe conflicts tended to show a higher blink rate. Ratings on family conflicts and anxiety over present illness showed this tendency more consistently over topics and periods than the ratings of sex conflicts. The correlations do not appear to be greater when the ratings in a given area are compared with the behavior on

TABLE 5
CORRELATIONS BETWEEN TEST SCORES, PSYCHIATRIC RATINGS, AND INTERVIEW BEHAVIOR

| Criterion | Eyeblink Rate | | | | | | | | | Verbal Rate | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Topics | | | | Periods | | | | | Topics | | | | Periods | | | | |
| | A | B | C | D | 1 | 2 | 3 | 4 | Total | A | B | C | D | 1 | 2 | 3 | 4 | Total |
| *Hy* | | | | | | | | | | −216 | | | | −229 | | | | |
| *Hi* | | | | | | | | | | −244 | −219 | | | −359** | | −221 | | |
| *D* | | 231 | 305* | 208 | | | 253 | | | −333** | | −251 | | −408*** | | −200 | | |
| *Pa* | | | | | | | | | | −223 | −287* | −203 | −206 | −426*** | | −237 | | |
| *Pt* | | | | | | | | | | | | | | −364** | | | | |
| *Sc* | | | | | | | | | | | | | | −328* | | | | |
| *Ma* | | | | | | | | | | | | | | | | | | |
| *Si* | | | 209 | | | | 205 | | | −310* | | | | −381** | −245 | | | −226 |
| Rating (Family) | 276 | 290* | 292* | 282* | 329* | 249 | 342** | 296* | 311* | | | | | | | | | |
| Rating (Sex) | 234 | 332** | 306* | | | 276 | 300* | 238 | 272 | | | | | 250 | | | 214 | |
| Rating (Illness) | 284* | | 309* | 303* | 356** | 312* | 242 | 219 | 284* | 271 | | | | | | 225 | | |
| MAS | | | | | | | | | | | | | | −260 | | | | |
| Cornell | | | | | | | | | | | | | | | | | | |

Note.—All correlations between −.20 and +.20 are omitted from this table. Decimal points omitted.
* $p < .10$.
** $p < .05$.
*** $p < .02$.

the topic in the same area, e.g., when ratings on family conflicts are compared with Topic B.

The psychiatric predictions of topics on which *S*s would be most and least uncomfortable were compared with the verbal rate and blink rate on these topics. Arcsin transformations were used to obtain *z* values for an estimate of the probabilities associated with these predictions (Walker & Lev, 1953). The *p* values represent the significance levels for the hypothesis of chance success ($p = .33$) in predicting Topics B, C, or D. Prediction of greatest discomfort on a topic coincided with *S*'s highest verbal rate on the same topic in 14 *S*s ($p = .231$). Lowest verbal rate coincided with prediction of least discomfort in 9 *S*s ($p = .14$ below expected chance value). *S*'s highest blink rate coincided with prediction of greatest discomfort in 16 *S*s ($p = .09$). Lowest blink rate coincided with prediction of least discomfort for that topic in 15 *S*s ($p = .13$).

## DISCUSSION

The results confirm the first general hypothesis, which stated that verbal rate is related to content. They also support the prediction that verbal rate increases as the emotional impact of a topic increases. The present results gain in significance when related to earlier findings (Kanfer, 1959) from similar interviews with a college population With college students, verbal rates among topics differed only when within-subject changes were considered and adjustment ratings were used to determine the topic on which the greatest momentary anxiety was expected to occur in the interview. In these *S*s, the topics did not seem to evoke similar emotional responses, and therefore rates varied over topics. In the present psychiatric population, the topic relating to hospitalization appeared to be uniformly anxiety-arousing in all *S*s. As the period means show (Table 1), verbal rate in Topic D started higher and stayed higher when compared with the other topics.

The lowest verbal rate was obtained on Topic C, dealing with *S*'s heterosexual attitudes. This finding is somewhat surprising, in view of the etiological importance which this topic might be assigned in the present population and the earlier findings of the high correlation between verbal rate and low adjustment on this topic with college students. A major reason for this result may be the different way in which the two populations handled this question. The late adolescent, unmarried students talked about dating, their sexual attitudes, and their conflicts and fears in this area. The older, married patients talked mostly about their husbands and generally expressed

positive feelings toward them. None of the patients discussed any sexual material. It is possible that the patients were evading any deep sexual problems. Presumably this would enable them to reduce their momentary anxiety in the interview.

The *family* topic (B), shows the second highest rate. This result is compatible with the findings of the highest correlation between verbal rate and adjustment ratings ($r = .467$) in the college student data.

The eyeblink data do not yield support for the second hypothesis. However, the data suggest that some common factor, such as the postulated momentary anxiety level, may partially underlie both behavioral changes even though these measures may not be highly correlated. In comparison with verbal rate, blink rate may be more sensitive, so that the special pertinence of unique events recalled during discussion of a topic may result in less uniformity over topics for all *S*s. The finding of a lower average blink rate on Topics A, B, and C certainly suggests that there was a general trend toward higher anxiety arousal on Topic D. The within-topic changes in blink rate parallel the pattern found on verbal rate. If momentary tension decreases as discussion of one topic continues, then both measures reflect this change in the predicted direction.

In the correlation matrix in Table 5, the size of the *r*'s must be evaluated in relation to the small *N*, which requires a high degree of relationship for significance. The distribution of *r*'s (Table 5) shows two distinct clusters of correlations. The self-inventory scores relate consistently to verbal rate during the initial 30-second periods. The psychiatric ratings, derived from behavioral observations and life histories, show a consistent, although low relationship to blink rate. In view of the moderate correlations between the variables reported here, due caution should be exercised by viewing these data as a basis for further research hypotheses; the correlations cannot be interpreted as validating relationships which have practical utility in clinical diagnosis.

The results suggest several post hoc speculations. The initially low verbal rate in MMPI high scorers may reflect a general inability to focus and organize their efforts at the start of a new "problem," i.e., a new topic. However, once a start is made, speech may come more easily under control of momentary tensions; hence, the correlations disappear. This is indicated by the decrease of correlations over periods within topics, and over topics as the interviews progressed.

The direction of the correlations between MMPI scores and verbal rate suggests that responsiveness in the interpersonal setting is greater in *S*s who show fewer specific symptoms or whose illness has not yet crystallized. Those *S*s who were most disturbed in their interpersonal relationships also tended to show the lowest verbal rate in the interview, possibly because they were least affected by *E*'s presence and unaware of any inappropriateness in their behavior. In this sense, their symptoms may well serve as efficient defenses in the interview. Although the dynamics of the depressive and the paranoid vary, the high correlations with verbal rate appear meaningful in both cases. The motor retardation in the depressed and his hypothesized repressed hostility towards others may reduce verbal rate as the severity of the illness increases. The paranoid patient's suspiciousness may similarly decrease his verbal output as symptoms increase.

Considering the frequent incongruence between psychiatric ratings and behavioral criteria, the cluster of correlations between rattings and eyeblink presents an interesting finding. It may be speculated that ratings by psychiatrists are often based on observation of somatic behaviors. Thus, the ratings would not be independent of *S*'s blink rate. However, such reasoning supports the pertinence of blink rate as a somatic concomitant of conflict behavior.

The overall results of this study provide further support for the existence of a relationship between verbal rate and verbal content in structured interviews. The data also suggest that during ongoing speech any verbal rate changes may relate to momentary tension level, to the extent that these changes are consistent with other components of the total patterning of *S*'s emotional state. The present study also demonstrates the utility of the experimentally structured interview in the evaluation of various dimensions of patient behavior. The method allows considerable control over *E*'s behavior without sacrificing the

richness or "natural" setting of the interview. In view of the rapidly growing evidence on the systematic effects of the interviewer's role on the patient's productions some controls over E's behavior seem essential in interview research.

## SUMMARY

The purpose of the present experiment was to investigate the relationship between verbal rate, content, and eyeblink rate in experimental interviews, and to provide further evidence for an earlier hypothesis that momentary anxiety increases verbal rate.

Thirty-six female psychiatric inpatients were given personality tests and were then seen for an experimental interview. The interview consisted of four questions to which the subject replied without interruptions for 2 minutes. The sessions were tape recorded and eyeblink rate was simultaneously recorded by an observer through a one-way mirror. The questions dealt with the following topics: (A) the present home, (B) family, (C) sexual attitudes, (D) reasons for hospitalization. Verbal rate was expected to be highest on Topic D, since this topic was selected a priori to be the most anxiety-arousing to the patient. In addition, each attending psychiatrist rated the patient with regard to degree of conflict in these areas.

Verbal rate and eyeblink rate were measured for 30-second intervals. MMPI scores, Taylor anxiety scores, and Cornell scores were also available for each *S*. Verbal rate was found to differ significantly over topics, with the highest rate on Topic D. Furthermore, there was a significant decline in speech rate within each topic, measured by comparing 30-second periods over all topics. The eyeblink data showed a similar decline within topics. While mean blink rate did not differ among topics, a significant topic × subjects interaction indicated that the subjects' blink rates were affected differently by the topics. A *t* test for correlated measures between average blink rate on Topics A, B, and C vs. D suggested that *S*s tended to yield higher blink rates on Topic D than on the average of the other topics.

Correlations between personality test scores, psychiatric ratings, and eyeblink and verbal rate revealed two clusters of moderate correlations. The first cluster contained significant correlations between high scores on five of eight MMPI scales and low verbal rate during the initial 30 seconds on each topic. The second cluster contained significant positive correlations between psychiatric ratings of degree of conflict and eyeblink measures.

The results supported the hypothesized relationship between verbal content and verbal rate. The eyeblink findings were considered as further evidence of the relationships between momentary tension, verbal rate and verbal content. The relationships of the dependent variables to the test scores and psychiatric ratings were discussed as suggestive of the importance of personality dynamics as determinants in the complex process of interpersonal communications.

## REFERENCES

DIBNER, A. S. Cue counting: A measure of anxiety in interviews. *J. consult. Psychol.*, 1956, **20**, 475–478.

DIBNER, A. S. Ambiguity and anxiety. *J. abnorm. soc. Psychol.*, 1958, **56**, 165–174.

DITTES, J. E. Extinction during psychotherapy of GSR accompanying "embarrassing" statements. *J. abnorm. soc. Psychol.*, 1957, **54**, 187–191.

DOEHRING, D. G. The relation between manifest anxiety and rate of eye blink in a stress situation. *USN Sch. Aviat. Med. res. Rep.*, 1957, No. 6. (Project NM 130199)

GOLDMAN-EISLER, FRIEDA. Speech-breathing activity: A measure of tension and affect during interviews. *Brit. J. Psychol.*, 1955, **46**, 53–63.

GOLDMAN-EISLER, FRIEDA. Speech-breathing activity and content in psychiatric interviews. *Brit. J. med. Psychol.*, 1956, **29**, 35–48.

KANFER, F. H. Effect of a warning signal preceding a noxious stimulus on verbal rate and heart rate. *J. exp. Psychol.*, 1958, **55**, 73–80.

KANFER, F. H. Verbal rate, content, and adjustment ratings in experimentally structured interviews. *J. abnorm. soc. Psychol.*, 1959, **58**, 305–311.

MAHL, G. F. Disturbances and silences in the patient's speech in psychotherapy. *J. abnorm. soc. Psychol.*, 1956, **53**, 1–5.

MEYER, D. R. On the interaction of simultaneous responses. *Psychol. Bull.*, 1953, **50**, 204–220.

MEYER, D. R., BAHRICK, H. P., & FITTS, P. M. Incentive, anxiety and the human blink rate. *J. exp. Psychol.*, 1953, **45**, 183–187.

PANEK, D. M., & MARTIN, B. The relationship between GSR and speech disturbances in psychotherapy. *J. abnorm. soc. Psychol.*, 1959, **58**, 402–405.

WALKER, HELEN M., & LEV, J. *Statistical inference.* New York: Holt, 1953.

(Received August 10, 1959)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 348-354

# THE SUSCEPTIBILITY OF CHRONIC SCHIZOPHRENICS TO SOCIAL INFLUENCE IN THE FORMATION OF PERCEPTUAL JUDGMENTS[1]

CARMI SCHOOLER[2] AND HERBERT E. SPOHN
*National Institute of Mental Health* *Veterans Administration Hospital, Montrose*

Lack of participation in social activity is such a prominent feature of the behavior of the chronic schizophrenic that it has led such diverse theorists as Cameron (1947), Fenichel (1945), and Fromm-Reichman (1939) to the general conclusions that the schizophrenic process involves a decline of responsiveness to social stimuli and that the degree of this impairment in schizophrenics varies with severity of illness.

The present study represents an attempt to test this general hypothesis experimentally. Hovland, Janis, and Kelley (1953) found that "persons who are most highly motivated to maintain their membership (in a group) tend to be most susceptible to influence by members of the group" (p. 276). If susceptibility to the influence of others is taken as an index of an interest in the social environment and responsiveness to social stimuli, it may be postulated that the degree of susceptibility to influence by others is inversely related to the degree of illness in chronic schizophrenics. The group situation first developed by Asch (1956) to study conformity behavior lent itself well to the experimental test of this hypothesis. In this procedure the subject (*S*) judges which of three comparison lines is the same length as a standard line after a group of the experimenter's confederates have given a wrong answer. The naive *S* is thus confronted by a choice between maintaining independence of judgment or conforming to the "norm" of the group. It was predicted that in the Asch situation regressed *S*s would conform less than partially remitted ones, who in turn would conform less than normal ones (Hypothesis I).

Two other variables were also investigated. One variable was introduced in order to study the possible effects on the hospitalized schizophrenic of the lack of contact with normals that characterizes his usual hospital environment. A previous study of the authors (Schooler & Spohn, in press) suggested that hospitalized schizophrenics accept degree of health as a prime criterion in selecting their social contacts and that their experience with the unreliability and inappropriateness of their peers' behavior also teaches them to distrust their peers' judgments. Consequently, it was predicted that schizophrenic *S*s to whom the confederates are introduced as normals would conform more than those to whom the confederates are introduced as fellow patients (Hypothesis II).

The third variable studied derives from the distinction made by Deutsch and Gerard (1955) between "normative" and "informative" sources of influence in the conformity situation. A person may conform in the Asch situation because he wishes to satisfy the positive expectation of the others (normative) or because he regards the others' judgments as superior to his own (informative). While in the standard Asch situation these two sources of motivation are implicit, they can, of course, be induced explicitly by instruction and thus function as two distinct additional sources of influence intensifying pressure to conform. No sound basis existed for predicting whether chronic schizophrenics respond selectively to these two sources of motivation to conform. It was of interest, however, to determine whether increments in pressure to conform would break through the hypothesized resistance to social influence in chronic schizophrenics. Specifically, it was predicted that chronic schizophrenics tend to conform more frequently when

[1] This report is based on a doctoral dissertation submitted by the senior author to the Graduate School of Arts and Sciences, New York University, February 1959. The senior author wishes to express his appreciation to Murray Horwitz for his guidance in the course of the study.

[2] This study was carried out while the senior author was a social psychology trainee at the Montrose Veterans Administration Hospital. The authors wish to express their gratitude to the staff of that hospital and to the staff of the Castle Point Veterans Administration TB Hospital. Special thanks are due to Jacob Cohen for his invaluable aid in dealing with the statistical problems which arose. A portion of this study was presented at the APA meeting in Washington, D. C. in August 1958.

normative or informative motivation to conform is explicitly induced than when this is not the case (Hypothesis III).

A series of hypotheses were also projected about the interactions among the several variables. These hypotheses were based on the reasoning that if degree of illness is inversely related to responsivity to social stimuli it would also be inversely related to readiness to respond differentially to relevant differences among social stimuli. Therefore, it was predicted that healthier *S*s tend to be more responsive to differences in the nature of the social pressure (Hypothesis IV-A) and the apparent mental health status of the confederates (Hypothesis IV-B).

## METHOD

### *Subjects*

One hundred and forty-four *S*'s were tested: 48 regressed schizophrenics, 48 partially remitted schizophrenics, and 48 hospitalized TB patients who served as normal controls. The regressed *S*s were chosen from the hospital's closed wards by clinical psychologists who used the Montrose Rating Scale (Rackow, Napoli, Klebanoff, & Schillinger, 1952–53); partially remitted *S*s were selected from the open wards through the same procedure. Only those *S*s whose score was below 10 on this scale were considered regressed and only those who were scored above 17 were considered partially remitted, these scores representing points on the scale that had been established previously as reliable cut-off points (Rackow et al., 1952–53). All of the schizophrenic *S*s had a history of at least 2 years of continuous hospitalization prior to the time of testing. The normal *S*s were chosen from among the patients of a nearby Veterans Administration TB hospital. None of those used had any known psychiatric history. All of the cells of the experimental design were equated for age and education (see Schooler, 1959). The mean age of the *S*s was 33.9 years; their mean level of education 10.2 years.

### *Design*

The schizophrenic *S*s were tested under a $2 \times 2 \times 3$ factorial design that permitted the following comparisons: partially remitted *S*s versus regressed *S*s, confederates introduced as normals versus confederates introduced as patients, standard Asch situation versus informative influence versus normative influence. Each of the 12 cells contained 8 replications. The normal controls were tested under the informative, normative, and standard Asch situations with 16 replications in each cell. For these *S*s the variable pertaining to mentally ill peer-confederate versus normal-confederate was not introduced on the grounds that the meaning of this variable would not be comparable. For TB patients the distinction between peers and nonpeers could only have been made on the basis of institutional status, while the major concern in manipulating the variable among the original population of schizophrenics was to isolate the effects of differences in the mental health status of the confederates.

### *Procedure*

The series of test cards, their order of presentation, and the confederates' responses duplicated Asch's procedure (1956). All *S*s were also pretested in an individual session several days before the experiment in order to insure that they were able to make all of the required comparisons correctly. Furthermore, schizophrenic *S*s were retested in an individual posttest immediately after the experimental situation.

In the actual experimental situation each *S* was tested with three confederates. The apparent mental health status of the confederates was varied by dressing the confederates who were to be presented to the schizophrenic *S*s as mentally ill patients in typical patient garb and introducing them as coming from wards housing patients of the same health level as the *S*. When the confederates were presented as normals, they wore the uniforms of their actual status—that of dietary worker.

The informative condition was implemented by asking each member of the group immediately before the test began, how well he did when he took the test earlier. Each of the confederates replied with a prearranged sentence which stated that he had made all of the comparisons correctly.

The normative condition was presented by having the confederates express approval of unanimity and disapproval of disagreement through a series of preprogramed comments, which were made the first three times the *S* agreed with the majority and the first three times he disagreed.

In the course of testing the normals it became apparent that many of the *S*s knew the confederates personally. Since this was not the case with the schizophrenic *S*s, it was necessary to gather information about the degree of acquaintance between the normal *S*s and the confederates in order to evaluate the effects of this difference. Consequently, the normal *S*s were asked to rate each of the others with whom they had taken the test on a 5-point scale:

> 1. never saw him before, 2. saw him around but do not know his name or know him enough to speak to, 3. know his name and know him well enough to say hello to, 4. have occasional conversations with him, and 5. he is a fairly good friend of mine.

### *Method of Analysis*

The original plan called for the use of analysis of variance. Since, however, the number of conforming errors was quite low (57% of the *S*s made no errors), the means of various cells were highly correlated with their variances (.98) and rank-order nonparametric statistics had to be employed. The Mann-Whitney *U* test was used for all two-condition analyses and the Kruskal-Wallis test for all three-condition analyses. Possible interactions between variables were explored

by a method developed by Jacob Cohen. In this method the error scores of *Ss* undergoing one condition of a given variable were randomly paired with the error scores of *Ss* undergoing the second condition of that variable, the other variables in the experiment being held constant. The differences between these paired scores were then ranked. These ranks were then summed separately for each of the conditions of the second variable and the sums compared through the Mann-Whitney or Kruskal-Wallis tests. Although this procedure does not represent a unique solution, since only one of many possible sets of random pairings was examined, the method was considered an adequate device for exploring interactions, especially since the fact that a large number of *Ss* produced no errors reduced the likelihood that a different random pairing would have produced different results.[3]

Several other problems arose, especially in the analyses of the data of the normal *Ss*. First of all, a correlation of .51 was found between the normal *Ss*' rating of acquaintance with the confederates and the amount of conformity shown. Consequently, it was decided that only those normal *Ss* would be included in the analyses whose degree of acquaintance with the confederates was similar to that existing between the schizophrenic *Ss* and the confederates with whom they were tested. Only those normal *Ss* were included who indicated by their scale responses that they did not know the names or have a speaking acquaintance with the confederates. Eight patients who met these criteria were randomly selected from each condition.[4]

Another problem that arose in analysing the responses of the normal *Ss* concerned the selection of that group of schizophrenic *Ss* who could appropriately be compared with the normals. Normal *Ss*, as noted above, could not be administered the treatment variation of peer-confederates versus normal-confederates since the association between mental health classification and status level differs for normals and schizophrenics. Since conclusive arguments could not be made for either alternative, comparisons with schizophrenics in both the peer-confederate and normal-confederate conditions were made. In addition, the responses of the schizophrenic *Ss* were analysed independently. Where no differences were found between the peer-confederate and normal-confederate conditions, the pooled data of all the schizophrenic *Ss* were also compared with those of the normals.

## RESULTS

### *Original Hypotheses*

Using the nonparametric methods described above, the three hypotheses concerning main effects[5] stemming from degree of mental health (I), status of confederates (II), and source of motivation (III) upon conforming behavior in chronic schizophrenics were tested. None of the differences in frequency of conforming responses between groups or conditions was found to be significant. Thus, the three major hypotheses of this study were not confirmed.

The hypotheses predicting second-order interactions resulting from differences between regressed and partially remitted schizophrenic *Ss* in selectivity of response to the several social discriminanda represented by the experimental conditions were similarly tested and not found significant. An exploratory analysis of the triple interaction between health, status, and nature of social pressure yielded an $H$ value of 5.890 (using the Kruskal-Wallis test) significant at the .05 level.[6]

As a result of the analysis of the triple interaction two tentative conclusions were reached. First of all, the variables studied seem to affect the number of conforming errors made only if they are introduced in a particular combination. Secondly, since the regressed *Ss* made proportionately more errors than did the partially remitted *Ss* in the normal-confederate informative condition as opposed to the peer-confederate informative condition, it seems that the effect of the informative condition is more dependent upon the mental health status of the majority among the regressed *Ss* than among the partially remitted ones. This difference suggests that compared with partially remitted patients, regressed patients give less weight to their peers as expert sources of information than they do to normals.

### *Nonpredicted Responses*

Asch (1956) assumed that all the errors made by his *Ss* reflected conformity to the responses of the majority. This assumption could not be made for all of the errors made by the schizophrenic *Ss*.

Besides the conforming errors, several types of responses occurred during the critical trials

[3] For a more complete description of the method and a description of the similar method through which triple interactions were analyzed see Schooler (1959).

[4] These new cells remained matched for age and education with the other cells (see Schooler, 1959).

[5] The data in both raw and ranked form are presented in Schooler (1959).

[6] The value of $H$ is significant at the .05 level of confidence if the beta approximation of the probability of a given value of $H$ occurring is used. It does not reach the .05 probability value if either chi square or alpha approximations are used.

which did not seem to directly reflect conformity to the responses of the majority. One such type occurred in those trials in which the third alternative line on the card, choice of which represented neither the correct answer nor the wrong answer given by the confederates, was not intermediate in length between the majority response and the correct answer. For example, in Trial 13, the correct choice was of a line 5 inches long, the majority chose a line 4 inches long, and the third alternative line was 6½ inches long. The choice of the third alternative in such a situation could not be construed to represent either an accurate response or a tendency to be influenced in the direction of the group. Such responses were labeled "noncorresponding" responses.

A second such type of error occurred in those trials in which the length of the third alternative line lay between the correct response and the one given by the confederates. Although Asch (1956) considered the choice of the intermediate alternative in such a situation a form of "partial conformity" to the norms of the group, the occurrence of the noncorresponding responses of the type just described casts some doubt on this assumption. It is conceivable that the choice of the intermediate line is based on the same motivation that led to the choice of the noncorresponding responses.

Two other types of errors also occurred besides those that occurred in the critical trials. One type, which was labeled "correct-wrong errors," occurred when the *S*s chose incorrect lines in those trials in which the confederates responded correctly. The other type occurred when *S*s made mistakes during the individually administered posttest.

This profusion of various types of errors raised the problem of combining the types into broader classes. One possible way of doing this was suggested by the fact that two distinct classes of errors could be distinguished on a priori grounds: the first class, nonconforming errors, including noncorresponding and correct-wrong errors, being made up of those errors which conform to neither the social pressure of the group nor the physical properties of the stimulus; the second class consisting of actual conforming errors.

In order to determine whether these two classes, distinguishable on a priori grounds, do in fact constitute two distinct empirical classes, the data of the regressed *S*s, the group that made the overwhelming number of nonconforming errors, were examined and intercorrelations among the various types of errors were obtained. The correlations shown in Table 1 support the assumption of two classes of errors. The correct-wrong and noncorresponding errors proved to be highly correlated with each other and not significantly correlated with the conforming errors. This finding seemed to justify a separate analysis for each of the classes.

TABLE 1
INTERCORRELATIONS AMONG TYPES OF ERRORS FOR THE REGRESSED SUBJECTS

| Type of Error | Correct-Wrong | Noncorresponding | Intermediate | Posttest | Conforming |
|---|---|---|---|---|---|
| Correct–Wrong | — | .85* | .70* | .63* | .27 |
| Noncorresponding | | — | .86* | .71* | .20 |
| Intermediate | | | — | .68* | .21 |
| Posttest | | | | — | .41* |
| Conforming | | | | | — |

* Significant at .01 level.

The table of intercorrelations also makes it readily apparent that the intermediate errors, at least among regressed schizophrenics, are empirically a part of the class of nonconforming errors. Thus, it would appear that at least for regressed schizophrenics, such errors do not represent partial conformity. However, since this type of error could not be categorized unambiguously on a priori grounds and might have a different meaning when made by *S*s in other health categories, its analysis was not combined with that of the nonconforming errors. A separate analysis was also performed for the posttest errors.

The analysis of the occurrence of nonconforming errors among the schizophrenic *S*s revealed that the occurrence of such errors is directly related to degree of illness ($z = 3.156$, $p < .01$) and the presence of conditions designed to increase the amount of pressure ($z = 2.313$, $p < .05$). Since the great majority of nonconforming errors was made by regressed *S*s, it would be fair to say that when such *S*s are subjected to increased pressure

towards conformity they respond by making a greater number of nonconforming errors.

Since in the analysis of the data from the schizophrenic *S*s neither the peer-confederate, normal-confederate variable, nor any of the interactions of which it was a part approached significance, it was possible to evaluate the data from the normal *S*s with the combined data from the schizophrenics from both the peer-confederate and normal-confederate conditions. This analysis showed that both degree of health ($H = 11.313$, $p < .01$) and the presence of conditions designed to increase the amount of social pressure applied ($z = 2.43$, $p < .02$) affected the number of nonconforming errors made. The difference between the health conditions was primarily due to the far greater number of nonconforming errors made by the regressed *S*s as opposed to the normal and partially remitted ones.

The analysis of the distribution of intermediate errors among the schizophrenic *S*s indicates that the occurrence of such errors is significantly affected only by degree of health ($z = 2.779$, $p < .01$). Since the peer-confederate–normal-confederate variable produced no significant main effects or interactions among the schizophrenics, it was possible to analyze the normal *S*s' data with the combined data of the schizophrenic *S*s from both the peer-confederate and normal-confederate conditions. This analysis showed that the production of intermediate errors was definitely related to degree of health, the regressed *S*s making many more such errors than the partially remitted or normal ones ($H = 11.030$, $p < .01$). However, no other significant differences were found.

The analysis of the posttest errors revealed that the occurrence of these errors was significantly related to degree of health, with the regressed *S*s making many more such errors than the partially remitted ones ($z = 2.645$, $p < .01$). A significant interaction was also found which indicated that more posttest errors were made after the normal-confederate informative condition than the peer-confederate informative condition ($H = 8.139$, $p < .02$).

## DISCUSSION

The results of this study do not confirm its basic hypothesis. Insofar as conforming behavior is an index of degree of responsiveness to the social environment, chronic schizophrenics are no less responsive to the influence of others than normals, nor are the severely ill less responsive than chronic schizophrenics who have attained some degree of recovery. The findings suggest that the manifest gross social withdrawal which characterizes many chronic schizophrenics *behaviorally* does not appear to be associated with a commensurate loss of sensitivity or awareness of others as social objects. Indeed, the results suggest that the hypotheses of this study should have been framed in reverse.

The significantly greater number of nonconforming errors made by the regressed *S*s suggests that as a group such *S*s are actually more reactive to the social stimuli of the Asch situation, albeit in an irregular way, than partially remitted and normal *S*s. Furthermore, both the effect of a particular combination of conditions in increasing the number of conforming errors made by the regressed *S*s and the increase of nonconforming errors under conditions designed to increase the amount of pressure suggest that such *S*s react differentially to the various interpersonal implications of the situations confronting them.

This possibility is of considerable theoretical interest since, by and large, as the hypotheses of this study suggest, most interpreters of chronic schizophrenia have postulated a diminution of social drives to account for the socially withdrawn behavior in chronic schizophrenics.

The nature of the motivation that underlies the nonconforming errors cannot be discerned from the data at hand. Two general hypotheses, consistent with the data, suggest themselves. On the basis of the high positive correlations between those types of errors (noncorresponding and correct-wrong) making up the class of nonconforming errors and the posttest errors, the type which seems most clearly to reflect disruption, one may postulate that the nonconforming errors occurred as part of a random pattern of responses reflecting disruption in the *S*s' cognitive functioning due to the stress of the experimental situation.

Such an explanation would be congruent with the conclusions of Rodnick and Gar-

mezy (1957). The findings they report indicate that

> although schizophrenic patients can and do respond adaptively in tasks of considerable difficulty and complexity . . . this adaptability is a tenuous one which can be disturbed by the introduction of minimal censure (p. 116).

It may be that the regressed patients saw the others with whose judgments they differed as potential sources of censure. This hypothesis is strengthened by the fact that more nonconforming errors occurred under the added pressure of the normative and informative conditions where disagreement may have been seen as more readily leading to censure.

On the other hand, it is possible that nonconforming errors represent the *S*s' explicit choice of an alternative which conforms to neither physical nor social reality. It is possible that the *S*s who chose such an alternative did so as a means of insulating themselves from the social environment of the test situation by abstaining from taking action that could be considered relevant by the others present. It can be hypothesized that this abstention functions to inform the others that the *S* cannot be judged according to the rules of the game because he is not a participant. Another hypothesis that views the nonconforming errors as the result of the *S*s' explicit choice of an alternative which conforms to neither physical nor social reality can be made if it is assumed that both agreeing with the majority and going against the majority to give the correct answer are psychologically painful. The nonconforming error may then represent an attempt to abstain from making either commitment. If either of the latter two hypotheses is correct, the posttest errors would represent a continuation of the policy of abstaining from making relevant choices.

A program of research aimed at differentiating among the above hypotheses is now under way. However, the study has direct implications even at present. Both the lack of any significant differences in the amount of conformity to social pressure shown by the *S*s of the different health categories tested and the presence of significant differences in the amount of inappropriate behavior shown by regressed schizophrenics under different conditions of social pressure emphasize the need for a careful re-examination of the nature of schizophrenic withdrawal and its relationship to degree of illness.

## SUMMARY

This study was designed to measure the effects of various factors possibly affecting the schizophrenics' readiness to be influenced socially. The basic technique used was the Asch situation. The hypotheses tested were based on the general theory that the schizophrenic process involves a progressive deterioration of responsiveness to the social environment. Thus, it was predicted that the amount of conformity shown would be inversely related to the degree of the subjects' illness.

Two other variables were also investigated. It was hypothesized that schizophrenic subjects to whom the confederates are introduced as normals would conform more than those to whom the confederates are introduced as mentally ill peers. It was also predicted that the amount of conformity exhibited would be increased by the presence of conditions directed at increasing the amount of normative and informative pressure.

Forty-eight regressed subjects, 48 partially remitted subjects, and 48 normal controls were tested. The major hypotheses were not supported by the experimental data. The variables under consideration only affected the amount of conformity shown in terms of a tentatively accepted triple interaction which indicated that for the regressed subjects the informative condition was more effective when presented by normal-confederates than by peer-confederates. In addition, some of the regressed subjects, although able to respond correctly on a pretest, produced a number of responses which conformed to neither the majority's choice nor the correct answer. Furthermore, the number of these responses was increased significantly under the normative and informative conditions, indicating that these regressed subjects are in fact responsive to the immediate social environment.

The results of the study were seen as contradicting the basic hypothesis that the schizophrenic process involves a progressive deterioration of responsiveness to the social environment. Instead of being less responsive than others to the immediate social environment, the re-

gressed subjects were more so, this increased responsiveness, however, taking the form of the production of responses which conform to neither physical nor social reality.

## REFERENCES


ASCH, S. D. Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychol. Monogr.*, 1956, **70**(9, Whole No. 416).

CAMERON, N. *The psychology of behavior disorders*. New York: Houghton Mifflin, 1947.

DEUTSCH, M., & GERARD, H. B. A study of normative and informational influence upon individual judgment. *J. abnorm. soc. Psychol.*, 1955, **51**, 629–636.

FENICHEL, O. The psychoanalytical theory of neurosis. New York: Norton, 1945.

FROMM-REICHMAN, FRIEDA. Transference problems in schizophrenia. *Psychoanal. Quart.*, 1939, **8**, 412–426.

HOVLAND C. I., JANIS, I. L., & KELLEY, H. H. *Communication and persuasion*. New Haven: Yale Univer. Press, 1953.

KRUSKAL, W. H., & WALLIS, W. A. Use of ranks in one-criterion variance analysis. *J. Amer. Statist. Ass.*, 1952, **47**, 583–621.

RACKOW, L. L., NAPOLI, P. J., KLEBANOFF, S. G., & SCHILLINGER, A. A. A group method for the rapid screening of chronic psychiatric patients. *Amer. J. Psychiat.*, 1952–53, **109**, 561–566.

RODNICK, E. H., & GARMEZY, N. An experimental approach to the study of schizophrenia. In M. R. Jones (Ed.), *Nebraska symposium on motivation*. Lincoln: Univer. Nebraska Press, 1957. Pp. 109–184.

SCHOOLER, C. The susceptibility to social influence of chronic schizophrenics in the formation of perceptual judgments. Unpublished doctoral dissertation, New York University, 1959.

SCHOOLER, C., & SPOHN, H. The social structure of a ward of chronic schizophrenics. *Int. J. soc. Psychiat.*, in press.


(Received August 10, 1959)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 355-359

# DIFFERENCES IN RESPONSE LATENCY AND RESPONSE VARIABILITY BETWEEN HIGH AND LOW ANXIETY SUBJECTS IN A FLICKER-FUSION TASK[1]

ROBERT A. WAGONER[2]
*Duke University*

THREE firm relationships have been demonstrated between anxiety level and performance in a series of psychophysical tasks. Goldstone (1953, 1955) and Jones (1955, 1958) have shown an inverse relationship between the level of manifest anxiety and measures of flicker-fusion threshold as well as a direct relationship between anxiety and several variability measures. A third index, the number of times a subject changes from one response category to another and back again (alternation of response), was found in a study by Wagoner and Cohen (1956) to be significantly larger for the highly anxious group. This finding was later confirmed by Jones (1955, 1958), who found reliable differences among high, medium, and low anxiety groups on this measure. These alternations may reflect an indecisiveness at the threshold which is exaggerated in the high anxiety group when very fine discriminations are required.

Careful study of some of the response characteristics related to variability around the threshold could shed light on the differences noted between high and low anxiety groups. In the situation confronting the subject in a flicker-fusion task, he is required to make a decision between either the category "flicker" or the category "steady" with regard to any one of the experimental stimuli. If discrete stimuli are presented in a sequence from low to high frequency (or the reverse), sufficient stimulus information is presumably present at the upper and lower extremes to produce consistent use of an appropriate category. In the vicinity of the threshold, the necessity for deciding which of the two available categories is appropriate may create a conflict situation. Cartwright's (1941a, 1941b) finding of heightened latency as a characteristic accompaniment of responses near the threshold suggests another method by which to assess events around the limen.

Typically, one would expect significantly larger response latencies for high anxiety subjects in the vicinity of their threshold than for low anxiety subjects. Unfortunately, this view provides no predictions with respect to the expected response latencies associated with the extremes of the continuum, those areas involving the consistent use of the "flicker" or the "steady" category.

One theoretical position that may help here is that of Taylor (1956) and Spence and Taylor (1951). Their position, derived from the Hullian notion of the interactive effect of drive and habit strength, suggests that a high drive state, such as anxiety, produces facilitation of performance where single or highly dominant response tendencies are present, but produces impairment of performance where multiple competing response tendencies are present. Multiple response tendencies are present in the vicinity of the threshold, and single, highly dominant response tendencies are found where subjects are responding consistently with either the "flicker" or the "steady" category. Consequently, mean latencies of high anxiety subjects should be higher than for low anxiety subjects within the threshold area, but lower mean latencies for the highly anxious subjects should be found where consistent "steady" or "flicker" responses are elicited.

## METHOD

*Subjects.* Twenty high anxiety outpatients (HA) and 20 low anxiety relatives (LA) of outpatients from Duke Hospital, closely matched for age and sex, were selected on the basis of two criteria. For the HA patients, a psychiatric diagnosis of anxiety reaction within a 6-month period was required; for the LA group, the absence of any history of psychiatric diffi-

[1] Based on a dissertation submitted to the Department of Psychology of Duke University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. Appreciation is gratefully expressed to L. D. Cohen for his supervision of this project.

This investigation was supported in part by Research Grant M887C from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service.

[2] Now at Veterans Administration Hospital, Indianapolis, Indiana.

culties was demanded. In addition, scores of 20 and above for the HA group and of 12 and below for the LA served as cutoffs on the Lorr Anxiety Rating Scale. The Lorr scale was always administered by the experimenter and one of two psychiatrists. Both raters had to agree as to placement above or below the cutoff point, or the potential subject was excluded from the study. Jones (1955, 1958) had previously shown that the Lorr Anxiety Rating Scale correlated .81 with the Taylor Manifest Anxiety Scale and .81 with the Duke Symptom Check List. Such high relationships made it possible to use only the Lorr Anxiety Rating Scale in conjunction with psychiatric diagnosis for the identification of groups.

*Equipment.* The apparatus employed in this study was similar to that used by Goldstone (1953, 1955). An electronically activated "glow modulator" tube (Sylvania R1131C) illuminated a 5-mm. circular test patch 30 cm. from the subject's eyes at a 1° visual angle. The stimulus, housed in a black box, appeared as a white light at a brightness of .97 millilamberts in a room dimly illuminated by a 60-w. lamp over the experimenter's recording papers. Binocular foveal fixation was used throughout. Each stimulus, at a light-dark ratio of 60:40 lasted 1.5 seconds. The onset of each stimulus activated a Standard Electric Timer calibrated in 1/100 second intervals. The clock stopped only when the subject either pushed a response lever, fixed in a comfortable armrest position, to indicate a judgment of "flicker" or pulled the lever to indicate "steady."

*Procedure.* After a partial dark adaptation period of about 5 minutes, each subject viewed a series of flickering lights presented in a modified method-of-limits procedure. A total performance consisted of three sets of trials interrupted by a 30-second rest interval. Each trial consisted of five ascending and five descending runs. Runs were terminated after three successive like responses, and the next run was begun one cycle above (ascending) or one below (descending) the terminal response of the preceding series. The percentage of reports of flicker for each half-cycle frequency for each trial separately and for all trials together were plotted on normal probability paper and a straight line of best visual fit was drawn by the experimenter. That number of cycles associated with the 50% point of flicker responses provided an estimate of the threshold, and the number of cycles difference between the 16% and the 84% points provided an estimate of the variability of response.

## RESULTS

A progressive decline in threshold values, independent of anxiety, complicates the calculations of the latency data. Table 1 shows the mean threshold values for the HA and LA groups for Trials I and III. For each group, the difference between the first and last trial is significant by $t$ test at the .01 level. Such downward shifts in threshold values during the experiment contraindicated the use of a single threshold for each subject for his entire performance. Rather, it required some procedure whereby the threshold for each run could be utilized. The adjustment adopted was based on

TABLE 1

MEAN THRESHOLD AND VARIABILITY MEASURES FOR 20 HIGH ANXIOUS AND 20 LOW ANXIOUS SUBJECTS

| | Lorr Scores Range | Mean Threshold | | | | Mean Variability (all trials) | Mean Alternations of Response |
|---|---|---|---|---|---|---|---|
| | | Mean Age | Total Trials | Trial I | Trial III | | |
| High Anxiety | 20–26 | 33.3 | 35.7 | 36.94 | 35.12 | 1.12 | 9.92 |
| Low Anxiety | 5–12 | 34.2 | 40.1 | 41.02 | 39.48 | .72 | 2.47 |

TABLE 2

HYPOTHETICAL RECORD OF SUBJECT'S RESPONSES TO A VISUAL DISCRIMINATION TASK

| Cycles | 30 | | 31 | | 32 | | 33 | | 34 | | 35 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Run | | | | | | | | | | | | |
| 1 | | | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | | |
| 2 | | | | | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| 3 | | | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | | |
| 4 | | | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | |
| Total | 5 | 5 | 5 | 5 | 5 | 4 | 3 | 1 | 0 | 0 | 0 | 0 |
| %0's | | | | | | 80 | 60 | 20 | | | | |

Note.—0 = flicker response; 1 = steady response.

TABLE 3

MEAN RESPONSE LATENCIES IN 1/100 SECONDS OF 20 HIGH ANXIOUS AND 20 LOW ANXIOUS GROUPS FOR THE THREE PORTIONS ON ALL TRIALS

| | Mean Latency F (100% Flicker) | Difference (T-F) | Mean Latency T (Threshold) | Difference (T-S) | Mean Latency S (100% Steady) | Difference (S-F) |
|---|---|---|---|---|---|---|
| High Anxiety | 107.2 | 93.0** | 200.2 | 46.3** | 153.9 | 46.7* |
| Low Anxiety | 119.2 | 43.5* | 162.7 | 36.9* | 125.8 | 6.6 |
| Differences HA-LA | −12.0* | | 32.5** | | 38.1* | |

* .05 level of confidence.
** .01 level of confidence.

an analysis of single runs, both ascending and descending.

As can be seen in Table 2, any run provides consistent flicker and consistent steady responses and may provide inconsistent alternations of response near the threshold. Thus, three segments of a run can be identified: a flicker portion (F), a steady portion (S), and a middle or threshold portion (T). To ensure an adequate sampling of latencies at the threshold where changes of response category occur, T includes not only latencies associated with alternations of response, which may be absent in some runs, but also those latencies associated with the last flicker response on an ascending run or the last steady response on a descending run. Arraying the data in accordance with the category of response, or portion, frees the response categories from their conventional dependence on the cycles per second at which they occurred and permits greater flexibility in the analysis of the data.

Table 1 shows that the HA group obtained a significantly lower threshold (35.7 cycles) than the LA subjects (40.1 cycles). The 4.4 cycle difference between these means is reliable at better than the .01 level. Mean variability is significantly higher for the HA subjects than for the LA group ($p < .01$). The HA group alternated between response categories more frequently than did the LA group ($p < .01$), by means of Wilcoxon's (1949) nonparametric $t$ test. Thus, in contrast to the LA group, the HA subjects generally displayed a reduced sensitivity to the flickering light, dispersed their judgments over a wider range of cycles, and shifted back and forth between response categories more frequently. These data are similar to the previous findings of Goldstone (1953, 1955) and Jones (1955, 1958).

Table 3 presents the mean response latencies associated with each of these portions of the run. Of particular importance is the observation that the HA group requires significantly ($p < .01$) more time to respond in T than does the LA group. Similarly, the HA subjects display a longer mean response latency than the LA subjects in their S ($p < .05$). In F, there is a reversal, and the HA group responds more promptly than the LA subjects ($p < .05$).

Within the LA group, no difference is noted between the mean latencies of their F and S. However, mean latencies of their S and F are both significantly different from those latencies accompanying responses in the T or threshold region ($p < .05$). For the LA subjects, T is associated with higher response latencies in comparison to latencies in their F and S. Like the LA group, the HA subjects require longer to respond in their threshold region than they do in their F ($p < .01$) or S ($p < .05$). Unlike the low anxiety subjects however, HA subjects respond more promptly in their F than in their S ($p < .05$).

## DISCUSSION

The latency results clearly indicate that the HA group requires more time to respond to threshold stimuli than do the LA subjects. These results support the general hypothesis that increased variability of response in the HA group is associated with increased difficulty in makir.g prompt and effective decisions at the threshold. While elevated relative to their own nonthreshold latencies, mean times at the threshold for the LA subjects did not reach the length of delay characteristic of the

HA group. Thus, the generally longer latencies of the HA group may indicate an exaggeration of a process found also in the LA group. Alternatively, these data may be explained with equal plausibility in terms of competing response tendencies. If mutually exclusive responses separately approach a maximal probability of occurrence at the threshold, increased time may be necessary for the most dominant one to emerge behaviorally in LA subjects. The presence of increased drive in the HA group could elevate the latencies, as observed in this experiment, above the level typical for the LA group.

Advancing the drive conception of anxiety as an alternative explanation of events occurring at the threshold presupposes the adequacy of the theory in dealing with responses not in competition. Since S and F consist of clearly dominant responses, the facilitative effects of heightened drive should reduce the mean of the HA group in these below that of the LA group. This decrement was observed only in F, where both groups were consistently using the flicker category. Contrary to expectation, the HA group took longer to respond in S than did the LA group. This reversal casts some doubt on the applicability of the drive conception of anxiety as applied to this experimental situation.

In this connection, one should point out that the task confronting the high and the low anxiety subjects is for each group somewhat different. Figure 1 shows the mean latencies of response arrayed around the common fixed threshold points (T). This fixed threshold, which was originally calculated to make it independent of cps, is plotted here to fall at the same number of cps as the conventional thresholds for the HA and LA groups. Only in the limited range of cycles falling between 35.0 and 40.0 were the two groups viewing a common series of stimuli. While the discrepancy in the exposed stimulus series perhaps did not interfere with the attempt to clarify events occurring at the threshold, different stimulus situations confronting the two groups may have contributed to some of the differences noted at the extremes of the stimulus series.

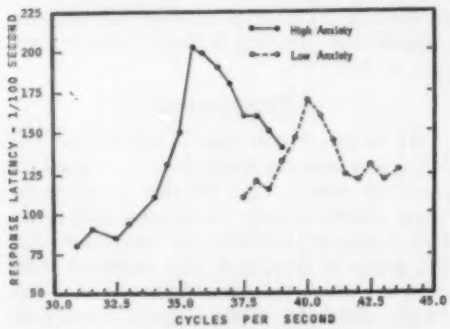


FIG. 1. Mean response latencies for each of the anxiety groups when arrayed around separate threshold points in terms of cycles per second.

The stimuli falling in the range between 35.0 and 40 cps are frequently noted to display a faint shimmering character (Gebhard & Mowbray, 1955). It is possible that the HA and LA groups adopt different criteria to deal with stimuli falling in this range. Different criteria may produce, on the one hand, sharply differentiated thresholds, while, on the other, they make it very difficult for the HA subjects to decide whether or not the stimuli in this range are steady or just faintly shimmering. If this is an adequate characterization of the stimulus situation confronting the HA group, sheer ambiguity of the stimulus series in this range may result in longer latencies. It is of some importance that, whatever these criteria may be, they are self-adopted and appear to be related to the variable of anxiety.

## SUMMARY

On the assumption that a subject's decision to use either of two available categories of response at the threshold in a flicker-fusion task is much like a conflict situation, HA subjects were expected to take more time to resolve the conflict and hence, display longer latencies at threshold than LA subjects. Where consistent responses were possible outside the threshold region, the drive conception of anxiety would predict that the HA subjects respond more promptly than LA subjects. These two hypotheses were tested in a flicker-fusion situation where latencies were recorded for each response. Twenty HA psychiatric outpatients and 20 LA relatives were matched for age and sex. The mean threshold latency of the HA group was considerably larger than that of the LA group, although the threshold latency of

each group was elevated above its own nonthreshold latency. Latencies accompanying nonthreshold responses were lower for the HA group only where subjects consistently responded with the category "flicker." Where consistent "steady" responses were possible, the LA group displayed shorter mean latencies than the HA group, a finding contrary to the hypothesis based on the drive conception of anxiety. This latter result may be a function of either the inappropriateness of the drive conception of anxiety to this task or an inherent stimulus ambiguity.

## REFERENCES


CARTWRIGHT, D. Decision-time in relation to differentiation of the phenomenal field. *Psychol. Rev.*, 1941, **48**, 425–442. (a)

CARTWRIGHT, D. The relation of decision-time to the categories of response. *Amer. J. Psychol.*, 1941, **54**, 177–196. (b)

GERHARD, J. W., & MOWBRAY, G. H. Differential sensitivity of the eye to intermittent white light. *Science*, 1955, **121**, 173–175.

GOLDSTONE, S. Flicker fusion measurements and anxiety level. Unpublished PhD thesis, Duke University, 1953.

GOLDSTONE, S. Flicker fusion measurements and anxiety level. *J. exp. Psychol.*, 1955, **49**, 200–202.

JONES, O. Relationship between visual and auditory discrimination and anxiety level. Unpublished PhD thesis, Duke University, 1955.

JONES, O. Relationship between visual and auditory discrimination and anxiety level. *J. gen. Psychol.*, 1958, **59**, 111–118.

SPENCE, K., & TAYLOR, JANET. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

TAYLOR, JANET. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **53**, 303–320.

WAGONER, R., & COHEN, L. D. Analysis of patterns of response of anxious and non-anxious subjects to a flickering light. *Percept. mot. Skills*, 1956, **6**, 167–170.

WILCOXON, F. *Some rapid approximate statistical procedures*. New York: American Cyanamid Co., 1949.


(Received July 20, 1960)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 359-364

# LEADER-MEMBER INTERACTION IN MANAGEMENT COMMITTEES

LAWRENCE SCHLESINGER,[1] JAY M. JACKSON,[2] AND JEAN BUTMAN
*University of Michigan*

STUDIES in leadership have shifted from a concern with characteristics and traits of leaders to a concern with the process of leadership, and with the exertion and acceptance of influence. As one step in this direction, the behavior of leaders has been reliably differentiated from that of members in analyses of the interaction processes in discussion groups. Kirscht, Lodahl, and Haire (1959) have recently summarized the existing data on categorical differences:

> . . . the general picture of a small group leader which appears common to various researches shows a group member who is . . . task-oriented, attempts to specify the problem, to suggest courses of action, to seek out the members' contributions, to integrate these and to propose solutions in the attempt to secure consensus in the group (p. 408).

At least one dimension, then, in which leader behavior differs from that of members may be described as the attempts of the leader to exert influence or control over the behavior of the members. Leadership acts may differ in the magnitude of attempted control as well as in frequencies of attempted control acts. The differences reported in frequency of attempted control behavior by perceived or designated leaders leaves open the question of the process whereby the behavior of the leader comes to be differentiated from that of the members. Do leaders of small groups actually make more influence attempts than the other members? Or do group members diminish their attempts to control the behavior of their fellow members?

Some evidence on these questions comes from a study in which the degree of control by the chairmen of existing management committees was manipulated by experimental appointment of chairmen who, in the opinion of their fellow members, varied in their directiveness. This experiment provided an opportunity to observe the influence process between chairmen and members under chairmen who differed in the degree of control they were perceived as exerting over group members.

## METHOD

*Subjects.* The study was conducted in a large utility company, on 23 existing committees of 106 middle-management executives. Each committee contained four or five members representing a major operating department and had the responsibility of coordinating interdepartmental problems in geographical districts of the company. The larger study[3] was designed to evaluate conditions influencing the performance of these interdepartmental management committees.

*Selection of experimental chairmen.* The independent variable in this study is the degree of control exercised by the designated chairmen of the committees. This manipulation was achieved without the knowledge of any of the participants including the chairman himself. Two months prior to the field experiment, all members of the committees responded to an extensive mailed questionnaire including a Chairman Behavior Questionnaire. In this questionnaire, they were asked to describe the typical behavior of each committee member when he was in the role of chairman, a position which rotated regularly among the members. The items used to elicit this description stem from the definition of "control" as a set of behaviors aimed at inducing restrictions on the behavior of the recipients. Scores derived from this questionnaire indicate the frequency with which the committee member in the role of chairman acts in the following way, according to other members of his committee: he does a great deal of guidance, he makes proposals for handling problems, he stimulates the group to action, he steers the discussion, he frequently gives his own opinions, and he often questions the statements of the members.

Using the mean of scores given by all other members of his group, then, a person was designated as a High Control or Low Control chairman depending upon whether his score fell above or below the median, both overall and in his own committee. These mean scores ranged, on a nine-point scale, from 7.56 to 2.17 with the median score at 5.04. We attempted to have the same number of departmental and divisional affiliations in each leadership condition.

*Procedure.* The team of investigators traveled to 23 different sites, where the interdepartmental committees usually met, but standardized all procedures and collected comparable data from each committee. Each committee worked on the solution of a simulated but realistic problem for 1.5 hours under controlled

[1] Now at 708 West Huron Street, Ann Arbor, Michigan.

[2] Now at University of Kansas.

[3] Institute for Social Research, Research Center for Group Dynamics. A study of Interdepartmental Management Committees. Unpublished manuscript, Ann Arbor, 1958.

conditions in the presence of three observers. The problem given to each committee was complex, involving considerable detailed information. It had been constructed by the investigators, with the help of a committee of executives, from problems typical of those faced by the committees in their regular meetings.

Each member was provided with an information sheet relevant to his department at the beginning of the meeting. The chairman was designated by the principal investigator, supposedly on a chance basis, and provided with additional instructions. The latter required that each group produce a detailed solution including recommended actions to implement their decision. The chairman was instructed to write up the minutes of the meeting, summarizing the decisions and recommendations reached during the discussion. The interest of higher management in the conduct and outcome of the study had been communicated to the executives, increasing the concern of the committees with their performance. Emphasis, however, was placed on the performance of the committee as a whole, and the executives were explicitly told that the study was not an assessment of individual abilities.

*Data collected.* The data relevant to this paper include observing and recording interactions during the meeting. Two management trainees of the company assigned to the study to act as observers recorded the interactions in each meeting with a system of categories based on Bales' (1950) system. The three relevant categories in the "control" dimension vary in magnitude of attempted control.

Control—statements aimed at restricting the behavior of the recipients. Three degrees of control are distinguished, depending on the latitude for compliance allowed the recipient:

1. Direction—the speaker orders or directs the other person or persons to whom the remark is addressed to do something giving the other little opportunity to refuse, e.g.: "We should read these first, then start the discussion."

2. Suggestion—the speaker suggests that the other do something, giving him room to refuse or make a countersuggestion, e.g.: "I suggest that Jack might start off by describing the situation in his department."

3. Request—the speaker asks the other to do something implying he has little right to ask and the other has every right to refuse, e.g.: "Would you be willing to tackle the problem of calculating the cost of the second alternative, or do you feel that can wait until we see the whole picture?"

## RESULTS

*Participant pre- and post-meeting ratings.* The original ratings of how the person behaved as chairman and ratings on the same items immediately after the experimental meeting, with that person as chairman, correlated .72, significant well beyond the .01 level for the 23 groups. The chairmen selected for the experimental conditions were seen to exhibit the same degree of control in the experimental meetings that the members had described in the questionnaire several months earlier.

*Participant and observer ratings.* The chairman's degree of control, as assessed by three neutral observers, was combined into a pooled rating that correlated .60 with the combined rating by participants, again significant well beyond the .01 level.

*Interactions of experimental chairmen.* Consistency of interaction recording between the two observers was measured by correlating the frequencies of interactions in a category attributed to a person during a meeting by each of the observers. The analysis was performed for a sample of five meetings. Where no interaction was attributed in a category by either observer, that pair of zero values was not included in the computation of the correlation, a decision which depressed the value of the correlations. As Table 1 indicates, the reliabilities thus measured show considerable variability, with only the "rejects" category reaching a less than acceptable level.

In describing the behavior of the chairman, two scores are possible: scores that describe the chairman's performance in relation to self, and ones that describe his performance in relation to the total group (Heyns & Lippitt, 1954). For each chairman there are two scores, one that expresses the concentration of verbal behavior in each category as a percentage of the chairman's own total behavior, and one that indicates what part of the total participation by the group, in each category, is accounted for by the chairman.

TABLE 1

CORRELATIONS BY INTERACTION CATEGORY BETWEEN OBSERVERS IN FREQUENCY OF TALLIES ATTRIBUTED TO SAMPLE OF INDIVIDUALS

| | Categories of Interaction | | | | | | |
|---|---|---|---|---|---|---|---|
| | Directs | Suggests | Requests | Accepts | Rejects | Gives Orientation | Asks for Orientation |
| *r* | .57* | .91** | .55* | .83** | .38 | .51* | .95** |
| *N* | 17 | 22 | 19 | 28 | 26 | 28 | 24 |

* Significant at .05 level.
** Significant at .01 level.

Looking first at the chairmen's specific acts in comparison to their own total behavior, in Table 2, there is no evidence that a chairman's degree of control score is significantly related to the percentage of his own behavior falling in any of the categories of interaction. Although the chairmen were reliably rated as exerting different degrees of control by both participants and observers, neither the magnitude nor frequency of control acts observed was related to rated degree of control. Since the percentage distributions may conceal the fact that High Control chairmen are more active in all categories than Low Control chairmen, a comparison was made of the absolute amount of activity of each of the two types of chairmen. The results were substantially the same, with no difference in the distributions or level of observed interactions. However, in Table 2, viewing the chairman's behavior in comparison with the other members of his group, i.e., as the percentage of the group's acts in various categories, the distinction between types of chairmen becomes clearer. In comparison to other members of their committee, High Control chairmen performed many more acts in the control dimension than Low Control chairmen, with highly significant correlations in both the suggests and requests categories. Viewing the behavior of the members alone, in the last column of Table 2, we find a negative relation between the percentage of the group's acts that are directive and the degree of control exercised by the chairman. Only the "requests" category reached significance.

TABLE 2

RELATIONSHIPS BETWEEN CHAIRMAN'S DEGREE OF CONTROL AND CONTROL BEHAVIOR OBSERVED IN EXPERIMENTAL MEETING

| Category of Interaction | Correlations[a] with Chairman Control | | |
|---|---|---|---|
| | Chairman's Behavior | | Member Behavior |
| | Percentage of Chairman's Acts | Percentage of Group's Acts | Percentage of Group's Acts |
| Directs | .16 | .28 | −.32 |
| Suggests | .15 | .51** | −.16 |
| Requests | .19 | .62** | −.41* |

[a] Pearson product-moment.
* Significant at the .05 level.
** Significant at the .01 level.

TABLE 3

RELATIONSHIPS BETWEEN MEMBER RATINGS ON DEGREE OF CONTROL EXERCISED AS CHAIRMEN AND OTHER CHARACTERISTICS

| Member Characteristics | Correlation[a] |
|---|---|
| Age | −.29 |
| Length of service | −.33 |
| Boss' influence on him | −.26 |
| His influence on boss | .09 |
| Ratings by others: | |
| Contribution to meetings | .56* |
| Skill as chairman | .75* |
| Friendly relations | .14 |

[a] Pearson product-moment.
* Significant at the .01 level.

*Differences among experimental chairmen.* To ascertain the significance to participants of the score for degree of control, the control score obtained from the Chairman Behavior Questionnaire was related to a number of other characteristics of the chairmen, and to ratings of them by the participants, obtained in a questionnaire filled out prior to the experimental sessions. The analysis was based on the entire sample of 106 members of the management committees.

In Table 3 it can be seen that High Control chairmen show some tendency, not statistically significant, to be younger, to have less service with the company and to feel somewhat freer of influence from their immediate superiors. The findings that reach a high level of significance refer to the way committee members rate each other on their skill as chairmen and on their general contribution to committee meetings. Members who are seen as high on control are rated as more skillful chairmen and also as contributing more to committee proceedings in their regular membership role. No relation was found between degree of control and feelings of personal friendliness among members. The results were substantially the same when the analysis was based on the 23 appointed chairmen.

## DISCUSSION

The problem of this paper is to account for the differences in interaction patterns of leaders and group members that have been reported in the literature. Conceptually, the elements of this experiment may be considered the goal of the committee members, the role of the chairman, the magnitude of control exerted by the

chairman over the members, the frequency of these attempted control acts by chairman and members, and the perceived resources of the chairman. First it seems reasonable to assume that as a result of the experimental instructions, and in the context of an evaluation of the committees, the goal of the committees was to provide a high quality solution to the problem presented. The role of the designated chairman may be conceived as a legitimate power role, following French and Snyder (1959) in which the incumbent has a legitimate right to influence the members and they an obligation to accept his influence. The two chairmen's interaction scores, based on the percentage of their own acts and the percentage of group acts in a category will be referred to as the individual and category scores, respectively, and have different conceptual meanings (Heyns, 1948). The individual score indicates the magnitude of control the person attempts to exert in response to the demands of the chairman role. The category score reflects the role the person plays in the group, including both his response to role demands and the acceptance of his influence by the members.

Considering first the chairmen's individual roles, the different chairmen evidently responded similarly to the demands on them in their role to direct and guide the meeting. Although both the participants and observers rated their behavior differently on the control dimension, they did not actually differ in the magnitude or frequency of control acts.

The category scores indicate that the High and Low Control chairmen had considerably different roles in the group. Compared to other group members, High Control chairmen exerted much more influence than did Low Control chairmen. Similar findings have been reported in Heyns' (1948) laboratory study of leader-member interaction patterns under chairmen instructed to either exercise or fail to exercise directive and guidance functions. Inspection of his data, which used identical scoring procedures and similar categories, indicates the greatest observed differences in controlling behaviors among leaders when their interactions are seen in relation to member behavior. The notion that leadership is not an attribute of a single agent, but a relationship among persons is strikingly supported.

The High Control chairmen, in their category scores, actually exerted a lesser magnitude of control, than did the Low Control chairmen. Their guidance implied relative autonomy for the members in comparison to directive control which leaves the members little choice but outright rejection or compliance. This combination of a higher frequency and lesser magnitude of control by the High Control chairmen suggests differences in response of the participants to the control acts of the chairmen.

When the participants do not perceive the designated leader as satisfactorily performing the controlling functions, the participants increase their own attempts to influence their fellow members. Heyns (1948) also reported that participants under the leader who was less controlling tended to perform more of the guidance functions than participants under the chairman who did exercise control. Carter, Haythorn, Shriver, and Lanzetta (1951) also reported that participants exerted more influence when the designated leaders did not respond directively. The greater frequency of influence attempts by participants in the Low Control groups did not crystallize into a greater emergence of informal leaders in these groups, as some industrial studies have reported. When the designated leader fails to perform adequately, considerable time is probably necessary for the members to decide which alternative persons they can confer leadership status upon.

The question remains of accounting for differences in participant response to the two kinds of leaders. Theoretically, leadership acts that are perceived as enabling the group members to reach their goal would be most acceptable to the participants. There is little doubt that the High Control chairmen were perceived as most capable of performing these leadership acts. The finding that degree of control was positively associated with two kinds of expertness, contributor to the group's tasks and skill as chairman, parallels Bales' (1953) report that the "best ideas" and "guidance" roles are closely associated in the participants' minds with leadership. As French and Snyder (1959) have reasserted, the effectiveness of an influence attempt by the leader increases with increasing perception that he is an expert in the area of the influence attempt.

It is noteworthy that leader acceptance, in this experiment, was based on task and chairmanship skills, and completely distinguished, as Bales (1953) found, from relations of interpersonal liking. As Gibb (1954) has suggested, the leader's influence upon group members is secondary to his influence upon goal locomotion. Although each of the chairmen attempted to exert the same amount of influence, the group members conferred leadership status upon those chairmen who were seen as providing a pathway to the group goal.

## SUMMARY

The problem of this study was to examine the influence process among leaders and members of small problem solving groups when the designated leaders vary in rated degree of control exerted. The opportunity for the study was provided by a field experiment on the degree of control exercised by the designated chairmen of management interdepartmental committees.

As part of an initial questionnaire, 106 members of 23 interdepartmental middle-management committees described, on a six-item questionnaire, the amount of control and regulation each member exercised in the role of chairman. A special meeting was then held in the field with half the committees under High Control, and half under Low Control chairmen, appointed "randomly" by the investigators. Each committee was observed for 1.5 hours while working on a simulated but typical company problem. The goal of the committees was to present the best possible solution of this problem to be assessed by higher management.

Observer and participant ratings of the chairman's behavior validated the original questionnaire ratings. But interaction profiles, based on the number of acts by a chairman in each category, showed no difference between the types of chairmen. They did differ strikingly from one another, however, when interaction scores were computed as ratios of the total number of acts performed by the entire committee in a particular category. Compared to other committee members, High Control chairmen were higher on the percentage of "control" acts performed. Committee members under Low Control chairmen made more attempts to influence group members. Analysis of the characteristics associated with the perceived degree of control exercised by the member when he was chairman indicated that the High Control chairman was more likely to be considered a skilled chairman and a more valuable contributor to the committee's work than the Low Control chairman.

It was concluded that the observed difference in leader-member interaction patterns is conditioned by the acceptance of the designated chairman as a resource by the group members. When the designated chairman is accepted as a resource capable of directing the group toward a successful solution of their task, he exerts a more frequent and lesser magnitude of control. The influence attempts of the Low Control leader are not accepted by group members, who vie with one another and the designated chairman for control over the group's activities. Differences in leader-member interaction patterns are thus seen to result from differences in the response of the members to leadership acts which vary in their perceived capacity to lead members to their goal.

## REFERENCES

BALES, R. F. *Interaction process analysis.* Cambridge: Addison-Wesley, 1950.

BALES, R. F. The equilibrium problem in small groups. In T. Parsons, R. F. Bales, & E. A. Shils (Eds.), *Working papers in the theory of action.* Glencoe, Ill.: Free Press, 1953. Pp. 111–161.

CARTER, L., HAYTHORN, W., SHRIVER, B., & LANZETTA, J. The behavior of leaders and other group members. *J. abnorm. soc. Psychol.*, 1951, **46**, 589–595.

FRENCH, J. R. P., JR., & SNYDER, R. Leadership and interpersonal power. In D. Cartwright (Ed.), *Studies in social power.* Ann Arbor: Institute for Social Research, 1959. Pp. 118–149.

GIBB, C. Leadership. In G. Lindzey (Ed.), *Handbook of social psychology.* Vol. II. Cambridge: Addison-Wesley, 1954. Pp. 877–920.

HEYNS, R. W. Effects of variation in leadership on participant behavior in discussion groups. Unpublished doctoral dissertation. Ann Arbor: University of Michigan, 1948.

HEYNS, R. W., & LIPPITT, R. Systematic observation techniques. In G. Lindzey (Ed.), *Handbook of social psychology.* Vol. I. Cambridge: Addison-Wesley, 1954. Pp. 370–404.

KIRSCHT, J. P., LODAHL, T. M., & HAIRE, M. Some factors in the selection of leaders by members of small groups. *J. abnorm. soc. Psychol.*, 1959, **58**, 406–408.

(Received August 11, 1959)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 365-369

# COMPETENCE AND CONFORMITY IN THE ACCEPTANCE OF INFLUENCE[1]

E. P. HOLLANDER

*School of International Service, American University*

WHEN one member influences others in his group it is often because he is competent in a focal group activity. A member may show such competence by individual actions that further the attainment of group goals (cf. Carter, 1954); more specific situational demands may variously favor the ascent of the expediter, advocate, or what Bales and Slater (1955) have termed the task specialist. An additional condition for the acceptance of influence involves the member's perceived adherence to the normative behaviors and attitudes of his group. His record of conformity to these expectancies serves to sustain eligibility of the sort Brown (1936) calls "membership character."

A person who exhibits both competence and conformity should eventually reach a threshold at which it becomes appropriate in the eyes of others for him to assert influence; and insofar as these assertions are accepted he emerges as a leader. But it is still necessary to account for the "nonconformity" that leaders display as they innovate and alter group norms. Certain shifts must therefore occur in the expectancies applicable to an individual as he proceeds from gaining status to maintaining it.

This process has been considered recently in a theoretical model of status emergence (Hollander, 1958). It features the prospect that behavior perceived to be nonconformity for one member may not be so perceived for another. Such differentiations are seen to be made as a function of status, conceived as an accumulation of positively disposed impressions termed "idiosyncrasy credits." A person gains credits, i.e., rises in status, by showing competence and by conforming to the expectancies applicable to him at the time. Eventually his credits allow him to nonconform with greater impunity.[2] Moreover, he is then subject to a new set of expectancies which direct the assertion of influence. Thus, whether for lack of motivation or misperception, his failure to take innovative action may cause him to lose status.[3]

It is readily predictable that in task oriented groups a member giving evidence of competence on the group task should with time gain in influence. If he simply nonconforms to the procedures agreed upon, the opposite effect should be observed. But the sequential relationship of nonconformity to competence is especially critical.

From the model, it should follow that, with a relatively constant level of manifest competence, the influence of a person who nonconforms *early* in the course of group interaction should be more drastically curtailed than in the case of a person who nonconforms *later*. Indeed, a reversal of effect would be predicted in the latter instance. Once a member has accumulated credits, his nonconformity to general procedure should serve as a confirming or signalizing feature of his status, thereby enhancing his influence. Accordingly, it may be hypothesized that given equivalent degrees of task competence, a member should achieve greater acceptance of his influence when he has conformed in the past and is now nonconforming than he should when nonconformity precedes conformity.

## METHOD

### *Design*

Twelve groups, each composed of four male subjects, were engaged in a task involving a sequence of

[1] This paper is based upon a study completed under ONR Contract 1849(00), while the author was at the Carnegie Institute of Technology. The views expressed here are those of the author and do not necessarily reflect those of the Department of the Navy.

The considerable assistance of H. Edwin Titus in this study is gratefully acknowledged.

Parts of the paper were reported at the symposium on "Recent Conceptions in Influence and Authority Process," held under the auspices of Division 8 at the 1959 APA Convention.

[2] This is a newer formulation of an observation long since made regarding the latitude provided leaders (e.g., Homans, 1950, p. 416). It is further elaborated in Hollander (1959).

[3] This proposition is consistent with various findings suggestive of the greater social perceptiveness of leaders (e.g., Chowdhry & Newcomb, 1952).

| | Green | Red | Blue | Yellow | Brown | Orange | Black |
|---|---|---|---|---|---|---|---|
| Able | −1 | −12 | +5 | −1 | −2 | +15 | −4 |
| Baker | +10 | −1 | −2 | −7 | +4 | −3 | −1 |
| Charlie | −5 | +5 | −3 | +3 | −11 | −1 | +12 |
| Dog | +5 | −7 | +10 | −2 | −5 | +1 | −2 |
| Easy | −4 | −1 | −1 | +1 | +13 | −10 | +2 |
| Fox | −6 | +15 | −5 | −1 | −3 | −1 | +1 |
| George | −1 | −1 | −2 | +10 | +4 | −2 | −8 |

FIG. 1. Matrix used in group task.

15 trials. A group choice was required for each trial from among the row alternatives in a 7 × 7 payoff matrix (see Figure 1). In every group, a fifth member was a confederate whose prearranged response was contrived to be correct on all but four trials, i.e., 2, 3, 6, and 12, thus reflecting considerable competence on the task. All interactions among participants took place through a system of microphones and headsets from partitioned booths. Subjects were assigned numbers from 1 to 5 for communicating with one another. The central manipulation was the confederate's nonconformity to procedures agreed upon by each group in a pretrial discussion. In terms of a division of the 15 trials into three zones—early, middle, and late—of 5 trials each, six treatments were applied: nonconformity throughout, nonconformity for the first two zones, for the first zone alone, for the last two zones, for the last zone alone, and a control with no nonconformity. In one set of treatments the confederate was designated number 5, and in the other number 4, to test possible position effects. Acceptance of the confederate's influence was measured by the number of trials by zone in which his recommended response was accepted as the group's. This was supplemented by post-interaction assessments.

### *Subjects*

The 48 subjects were all juniors in the College of Engineering and Science at the Carnegie Institute of Technology. All had volunteered from introductory psychology sections after being told only that they would be taking part in a study of problem solving in groups. Care was taken in composing the 12 groups so as to avoid either placing acquaintances together or having membership known in advance. Thus, no two subjects from the same class section were used in the same group, and subjects reported at staggered times to different rooms. By the time a subject reached the laboratory room where the experiment was actually conducted, he had been kept apart from the others and was not aware of their identity. The subjects never saw one another during the entire procedure, nor were their names ever used among them.

### *Instructions and Set*

Once seated and assigned a number, every subject was given a sheet of instructions and the matrix used for the task. These instructions fell into two parts, both of which were reviewed aloud with each subject individually, and then with the entire group over the communication network. The first part cautioned the subjects to always identify themselves by number (e.g., "This is Station 3 . . . ") before speaking and not to use names or other self-identifying references. The second part acquainted them with the procedures to be used, emphasized the aspect of competition against a "system," and established the basis for evident procedural norms. It read as follows:

1. You will be working with others on a problem involving a matrix of plus and minus values. Everyone has the same matrix before him. The goal is to amass as many plus units as possible, and to avoid minus units. Units are worth 1 cent each to the group; the group begins with a credit of 200 units. You cannot lose your own money, therefore. There will be fifteen trials in all.

2. In any one trial, the task involved is for the group to agree on just *one* row—identified by Able, Baker, Charlie, etc.—which seems to have strategic value. Once the group has determined a row, the experimenter will announce the column color which comes up on that trial. The intersecting cells indicate the payoff. Following this announcement, there will be thirty seconds of silence during which group members can think individually about the best strategy for the next trial, in terms of their notion about the system; note please that there are several approximations to the system, although the equation underlying it is quite complex. But work at it.

3. At the beginning of each trial the group members must report, one at a time, in some order, as to what they think would be the best row choice on the upcoming trial. Members may "pass" until the third time around, but must announce a choice then. Following this, groups will have three minutes on each trial to discuss choices and reach some agreement; this can be a simple majority, or unanimous decision; it is up to the group to decide. If a decision is not reached in three minutes, the group loses 5 units.

4. Before beginning the trials, the group will have five minutes to discuss these points: (*a*) The order of reporting; (*b*) How to determine the group choice for a given trial; (*c*) How to divide up the money at the end. These decisions are always subject to change, if the group has time and can agree. After the 15th trial, group members may have as much as five minutes to settle any outstanding decisions. Then headsets are to be removed, but group members remain seated for further instructions, and the individual payment of funds.

### *Instruments and Procedure*

The matrix was specially constructed for this study to present an ambiguous but plausible task in which

alternatives were only marginally discrete from one another.[4] The number of columns and rows was selected to enlarge the range of possibilities beyond the number of group members, while still retaining comprehensibility. The fact that the rows are unequal in algebraic sum appears to be less important as a feature in choice than the number and magnitude of positive and negative values in each; there is moreover the complicating feature of processing the outcome of the last trials in evaluating the choice for the next. All considered, the matrix was admirably suited to the requirements for ambiguity, challenge, conflict, immediate reinforcement, and ready manipulation by the experimenter.

The confederate, operating as either 4 or 5 in the groups, suggested a choice that differed trial by trial from those offered by other members; this was prearranged but subject to modification as required. Since subjects rather typically perceived alternatives differently, his behavior was not unusual, especially during the early trials. For the 11 trials in which the confederate's row choice was "correct," the color that "came up" was contrived to yield a high plus value without at the same time providing a similar value for intersection with another person's row choice. Had his recommendation been followed by the group on these trials, high payoffs would have accrued.

The device of a 5-minute pretrial discussion had special utility for establishing common group expectancies, in the form of procedures, from which the confederate could deviate when called for in the design. Predictable decisions on these matters were reached unfailingly. But their importance lay in having a *public affirmation* of member intent. Thus, on order of reporting, it was quickly agreed to follow the order of the numbers assigned members. Each group, despite minor variants suggested, decided on simple majority rule. Regarding division of funds, equal sharing prevailed, sometimes with the proviso that the issue be taken up again at the end.

In the zones calling for nonconformity, the confederate violated these procedures by speaking out of prescribed turn, by questioning the utility of majority rule, and by unsupported—but not harsh—challenges to the recommendations made by others. He manifested such behaviors on an approximate frequency of at least one of these per trial with a mean of two per trial considered optimum. Thus, he would break in with his choice immediately after an earlier respondent had spoken and before the next in sequence could do so; when there were periods of silence during a trial he would observe aloud that maybe majority rule did not work so well; and he would show a lack of enthusiasm for the choice offered by various others on the matter of basis. Lest he lose credibility and become a caricature, in all instances he chose his moments with care and retained an evident spontaneity of expression.[5]

[4] The matrix is an adaptation, at least in spirit, of a smaller one used with success by Moore and Berkowitz (1956).

[5] The same person, H. E. Titus, was the confederate throughout.

## RESULTS AND DISCUSSION

The task gave quite satisfactory signs of engrossing the subjects. There was much talk about the "system" and a good deal of delving into its basis, possibly made the more so by the subjects' academic background; the returned matrices were littered with diagrams, notations, and calculations. Though quite meaningless in fact, the confederate's tentative accounts of his "reasoning" were evidently treated with seriousness, perhaps as much because of the contrived time constraint, which prevented probing, as of his jargon regarding "rotations" and "block shifts." In any case, the confederate at no time claimed to have the system completely in hand. He delayed his response from the sixth trial onward to suggest calculation of an optimum choice in the face of conflicting alternatives; and the four trials on which he was "wrong" were spaced to signify progressive improvement, but not total perfection.

Most pertinent, however, is the fact that there were no manifestations of suspicion concerning the confederate's authenticity. The others seemed to believe that he was one of them and that he was "cracking" the system; the post-interaction data were in full agreement.

Since all of the interactions were available on tape, it was possible to derive a number of indices of acceptance of influence. The most broadly revealing of these appeared to be the frequency of trials on which the confederate's recommended solution was followed.

In Table 1 this index is employed to discern the effects of three major variables. The analysis is arranged by zones (Z) of trials, and in terms of the confederate's nonconformity (NC) in the *current* zone and immediate *past* zone.[6] The means given in each cell indicate the number of trials, out of five per zone, on which the confederate's choice was also the group's. In a chi square test, the effect of position upon this measure was found to be nonsignificant, and is therefore omitted as a distinction in the analysis of variance.

[6] For Zone I, the "past zone" refers to the discussion period. If he was to nonconform there, the confederate would question majority rule and suggest that the division of funds be left until the end rather than agree then on equal shares.

TABLE 1

MEAN NUMBER OF TRIALS ON WHICH A GROUP ACCEPTS CONFEDERATE'S RECOMMENDED SOLUTION

| Confederate's Previous Conformity | Zone I (Trials 1–5) | | Zone II (Trials 6–10) | | Zone III (Trials 11–15) | |
|---|---|---|---|---|---|---|
| | Nonconforming[a] | Conforming | Nonconforming | Conforming | Nonconforming | Conforming |
| *With* Procedural nonconformity in immediate *past* zone | 1.67<br>6[b] | — | 3.25<br>4 | 3.00<br>2 | 4.00<br>4 | 5.00<br>2 |
| *Without* Procedural nonconformity in immediate *past* zone | — | 2.00<br>6 | 5.00<br>2 | 3.75<br>4 | 5.00<br>2 | 4.75<br>4 |

ANALYSIS OF VARIANCE

| Source | *SS* | *df* | *MS* | *F* |
|---|---|---|---|---|
| Current Nonconformity | .20 | 1 | .200 | — |
| Zones | 47.05 | 2 | 23.525 | 35.01** |
| Past Nonconformity | 3.36 | 1 | 3.360 | 5.00* |
| Int: Current NC × Z | 1.22 | 2 | .610 | — |
| Int: Current NC × Past NC | 13.52 | 1 | 13.520 | 20.12** |
| Int: Z × Past NC | .72 | 2 | .360 | — |
| Int: Current NC × Z × Past NC | 4.11 | 2 | 2.055 | 3.06 |
| Residual | 16.12 | 24 | .672 | |
| Total | 86.30 | 35 | | |

[a] Confederate showed procedural nonconformity on the trials in this zone.
[b] Indicates number of groups upon which cell is based.
* $p < .05$.
** $p < .001$.

The significant *F* secured from Zones is in accord with prediction. It reveals the ongoing effect of task competence in increasing the acceptance of the confederate's choice, to be seen in the rising means across zones. While current nonconformity does not yield a significant effect, past nonconformity does. Viewing the table horizontally, one finds that the means for "without" *past* NC exceed the means for "with" *past* NC in all instances but one. Regarding the significant interaction of *current* and *past* NC, the combination "without-without" has a sequence (2.00, 3.75, 4.75) of persistently higher value than has "with-with" (1.67, 3.25, 4.00); this, too, is in line with prediction. Finally, the maximum value of 5.00 in Zone II for the combination "without" *past* NC but "with" *current* NC confirms the key prediction from the model, at least within the context of the relative magnitudes there; the same value is also seen in Zone III for the identical combination; still another reading of 5.00 holds there, however, for the inverse combination, but in a tight range of values quite beyond separation of effects for interpretation.

Considerable consistency was found too in the post-interaction data. On the item "overall contribution to the group activity," 44 of the 48 subjects ranked the confederate first; on the item "influence over the group's decisions," 45 of the 48 ranked him first. Two things bear emphasis in this regard: subjects had to individually write in the numbers of group members next to rank, hence demanding recall; and their polarity of response cut across all six treatments, despite significant differences among these in the actual *acceptance of influence*. That the confederate therefore made an impact is clear; but that it had selective consequences depending upon the timing of his nonconformity is equally clear.

In detail, then, the findings are in keeping with the predictions made from the model. The operational variable for measuring acceptance of influence was confined to the task itself, but nontask elements are touched as well. In that respect, the findings corroborate the subtle

development of differential impressions as a function of even limited interpersonal behavior.

Some unquantified but clearly suggestive data are worth mentioning in this regard. Where, for example, the confederate began nonconforming *after* the first zone, his behavior was accepted with minimal challenge; by the third zone, his suggestion that majority rule was faulty yielded a rubber stamping of his choice. Again, if he had already accrued credit, his pattern of interrupting people out of turn not only went unhindered but was taken up by some others. Quite different effects were elicited if the confederate exhibited nonconformity from *the outset*, notably such comments of censure as "That's not the way we agreed to do it, five."

The findings are especially indicative of the stochastic element of social interaction and its consequence for changing perception. Especially interesting is the fact that these effects are produced even in a relatively brief span of time.

## SUMMARY

A study was conducted to test the relationship between competence on a group task and conformity or nonconformity to procedural norms in determining a person's ability to influence other group members. Data were gathered from 12 groups engaged in a problem solving task under controlled conditions. Each was made up of five members one of whom was a confederate who evidenced a high degree of competence during the 15 trials. His nonconformity to the procedural norms agreed upon by the group was introduced at various times, early, middle, or late, in the sequence of trials. Influence was measured by the number of trials (per segment of the entire sequence) in which the confederate's recommended solution was accepted as the group's choice. As a broad effect, it was found that a significant increase in his influence occurred as the trials progressed, presumably as a function of the successive evidences of competence. Past conformity by the confederate was also found to be positively and significantly related to the acceptance of his influence; finally, there was a statistically significant interaction between past and current nonconformity reflected in high influence in the groups in which the confederate had conformed earlier in the sequence of trials but was presently nonconforming. These results were all thoroughly consistent with predictions made from the "idiosyncrasy credit" model of conformity and status.

## REFERENCES

BALES, R. F., & SLATER, P. E. Role differentiation in small decision-making groups. In T. Parsons, R. F. Bales, et al. (Eds.), *Family, socialization, and interaction process*. Glencoe, Ill.: Free Press, 1955.

BROWN, J. F. *Psychology and the social order*. New York: McGraw-Hill, 1936.

CARTER, L. F. Recording and evaluating the performance of individuals as members of small groups. *Personnel Psychol.*, 1954, **7**, 477–484.

CHOWDHRY, KAMLA, & NEWCOMB, T. M. The relative abilities of leaders and non-leaders to estimate opinions of their own groups. *J. abnorm. soc. Psychol.*, 1952, **47**, 51–57.

HOLLANDER, E. P. Conformity, status, and idiosyncrasy credit. *Psychol. Rev.*, 1958, **65**, 117–127.

HOLLANDER, E. P. Some points of reinterpretation regarding social conformity. *Soc. Rev.*, 1959, **7**, 159–168.

HOMANS, G. C. *The human group*. New York: Harcourt, Brace, 1950.

MOORE, O. K., & BERKOWITZ, M. I. Problem solving and social interaction. *ONR tech. Rep.*, 1956, No. 1. (Contract Nonr-609(16), Yale University Department of Sociology)

(Received August 17, 1959)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 2, 370-379

# CREATIVITY AND PRIMARY PROCESS:

## A STUDY OF ADAPTIVE REGRESSION[1]

FRED PINE AND ROBERT R. HOLT

*Research Center for Mental Health, New York University*

BEGINNING with Freud (1921, 1953), but especially since Kris' explorations of creative processes (Kris, 1952; also Pappenheim & Kris, 1946), psychoanalytic theorists have described relationships among creativity, primary process (Freud, 1955), and a particular kind of ego control that permits adaptive use of primary process. The concept "regression in the service of the ego" (Kris, 1952; see also Schafer, 1958) refers to a momentary and at least partially controlled use of primitive, nonlogical, and drive dominated modes of thinking in the early stages of the creative process. The present study is one of the degree to which characteristic individual modes of expression and control of primary process in the Rorschach test are related to the quality of productions created in a variety of experimental tasks.

The study of creativity can begin from at least three different standpoints: the process, the product, and the person. Kris' work centers primarily though by no means entirely on the creative *process*, on the dynamic events surrounding the creative act itself. At the moment of creation, thinking is more fluid and often more drive dominated and less logical than usual; it is generally pervaded by intense emotionality—feelings of omnipotence, intense pleasure and/or pain (cf. Milner, 1957). Thinking closer to the primary process type is thus involved, but in the service of a productive act geared eventually to the creation of a communicable and socially valued product. The oft-described phases of creativity (cf. Kris, 1952; Rapaport, 1951) reflect this total process: the more primitive thinking characterizes the "inspirational" phase while the more controlled thinking characterizes the "elaborative" phase.

In a recent research, Pine (1959) studied the *products* of creative activity, employing the concepts of drive expression and drive control. The quality of productions—of essentially uncreative subjects (*S*s)—in a literary task (TAT) and in a task of scientific hypothesis formation varied with the amount and control of drive content in the product. Not only did the better productions include more drive content, but in addition such content was better controlled than in the poorer productions. Controls were reflected in the effective integration of the drive content into the total created production as well as in modulation of drive intensity and socialization of drive aim.

By contrast, the present study approaches the problem of creativity—and uncreativity—from the standpoint of the *person*, the creator. It involves a number of assumptions, from which three hypotheses were generated. We assume, first, that it is possible to score quantitatively evidences of the primary process that are apparent in a finished thought product—in the present case, in Rorschach test responses. Further, we assume that people differ in the degree to which they permit their responses to the Rorschach test to manifest recognizable signs of the primary process, or are unable to prevent its emergence in their responses (processes which we call, for short, the expression of primary process) and in the effectiveness of their control over such expressed primary process content. Finally, we assume that a score taking into account both the amount of expression of primary process and the effectiveness with which it is controlled constitutes an operational measure of adaptive regression (Kris' regression in the service of the ego). What we cannot assume, however, and what this study undertakes to explore, is the extent to which styles of expression and control of primary process on the Rorschach test are generalized and apply to a variety of tasks not all of which seem to require primary process modes of thinking.

Hypothesis 1. *S*'s tendency to express more or less primary process is unrelated to the quality of the productions he creates in various tests of imagination.

[1] This research was supported by a grant from the Ford Foundation.

Hypothesis 2. *S*'s tendency towards effective control over expressed primary process material is positively related to the quality of his created productions, effective control being associated with higher quality.

Hypothesis 3. *S*'s tendency towards adaptive or maladaptive regression to primary process modes of thinking (based on an evaluation of the combined expression and control tendencies) is positively related to the quality of his created productions, adaptive regression being associated with higher quality.

These hypotheses stem from an assessment of the relative significance of control—in contrast to expression—of primary process for creativity. Amount of expression per se takes no account of the control processes required for creative work. It is expected that *S*s who produce a good deal of primary process material on the Rorschach test with consistently good control will give the highest quality productions in the tests of imagination (Hypothesis 3). It is expected, further, that individual differences in control over primary process expression will be related to the quality measures, even apart from consideration of the amount of expression (Hypothesis 2). This is because almost all *S*s express at least some primary process content on the Rorschach test, because the tests of imagination that were used by and large do not require a capacity for expressive thinking comparable to that required for the production of, say, a major work of literature, and because we did not expect our sample of college students to contain many persons who would regress pathologically.

In testing these hypotheses, the study has one additional aim: to provide validation for Holt's system for scoring expression and control of primary process material in the Rorschach test (Holt, 1959; Holt & Havel, 1960). Previously, this system has been shown to provide an effective predictor of response to perceptual isolation, particularly in terms of the operation of primary process thinking during isolation procedures (Goldberger, 1958; Goldberger & Holt, 1961).

## METHOD

*Subjects.* *S*s were 13 male and 14 female undergraduates at a large urban university. They were a group of volunteers who were intensively studied through clinical and experimental research techniques by numerous investigators for about one year. *S*s had been selected for the research program on the basis of (*a*) approximate indications on the MMPI of emotional stability, and (*b*) intelligence (upper 60% of entering college freshmen) on the Ohio State Psychological Examination. They have no particular excellence in creative arts or science nor, with two exceptions, do they have any professional commitment to these areas. *S*s are the same as those used in Pine's (1959) related study. The present study focuses on seven tests administered to all *S*s. "Creativity" scores were derived from each test, but the nature of this score varies with the particular test; in aggregate they are referred to as the "quality" scores.

*Thematic Apperception Test (TAT).* The TAT was used in investigating the quality of literary productions. The test was administered individually. Standard instructions were given and *S*s were told to take about 5 minutes per story and, on Card 11, to "tell a fairy tale such as you might tell a child." Nine stories per *S* were considered here, those given to Cards 1, 4, 11, 13MF, 16, 2, 18GF, 19, and 18BM (males) or 15 (females). Stories were recorded verbatim, the first five tape recorded and the last four written out by *S*.

The stories were rated on 12 attributes that were expected on a priori grounds to reflect the *literary quality* of the story: length, figures of speech, vocabulary, variety in sentence structure, naming characters, use of direct speech, characterization, time perspective, unity and continuity, genuineness, originality, and specification.[2] Fuller descriptions of these variables are given elsewhere (Pine, 1959); an attempt was made to control for length as much as possible in the latter 11 variables. Each of these story attributes was rated by one rater; five raters were used in all.[3] Although it was necessary to use each rater for more than one rating, only ratings which were unlikely to be confused with one another (e.g., variety in sentence structure and

[2] We wish to acknowledge our indebtedness to D. J. van Lennep, whose work in this area suggested some of these variables for analysis.

[3] One of the authors (FP) and four other raters carried out these ratings. Only FP knew anything about *S*s' scores on the other creativity tests, and he rated only the objective variables on the TAT (length, number of stories with direct speech, number of stories in which characters are given names). One of the TAT raters had some share in the ratings of the independent Rorschach variables. Although this rater had only a part share in rating the TAT (rating 2 of the 12 variables) and only a part share in rating some of the Rorschach protocols, correlations between the TAT quality scores and the Rorschach independent variables for handling of primary process were carried out with her ratings of the TAT excluded from the TAT score to check on any possibility of rating contamination. The correlations obtained did not differ in pattern or in statistical significance from those described below (Table 1) for the total TAT score.

We wish to express our acknowledgment to the many members of the staff of the Research Center for Mental Health who carried out various rating tasks.

characterization) were assigned to one rater. *Ss* were ranked on each of the 12 story attributes and final ranks were derived by summing the 12 for each *S* and re-ranking the totals. The final combined ranking was used in all data analyses. The scoring reliability of the final literary quality scores was determined by the intraclass correlation technique (Ebel, 1951) applied to the 12 rankings; reliability of the mean score was .82 for the total sample.

*The Science Test.* The Science Test consisted of two problems, each posing a scientific dilemma which *S* was asked to resolve. One item (#2) was drive related and was generally felt by *Ss* to be far the more fanciful while the other (#1) was more neutral. *Ss* were required to write out three theories to account for each hypothetical situation. Instructions were to make the theories complete and realistic (i.e., probable) and to use imagination in a down-to-earth scientific manner. There was no fixed time limit, but no *S* took longer than one hour. The test called for skill in forming imaginative hypotheses; analysis and synthesis of information were only minimally involved and hypothesis testing was not involved. The problems were as follows:

1. A general idea of the way that coal forms is first that trees fall and decay; then owing to certain upheavals of the earth, land gets on top of them and over time the pressure of the earth turns the original material into coal. There are no trees in Antarctica, yet I want you to assume that coal has been found there. How could this be accounted for?

2. It is known that the differentiation of male and female sexes came about fairly early in animal evolution. That is, even in the lower animals male and female sex organs are in different individuals—there are separate male and female sexes. I want you to assume that recently a very advanced animal species, almost like the human species, has been found in which all individuals have both male and female sex organs. How could this be accounted for?

Ratings for *quality of theories* (on a four-point scale) were based on the specific theories given by *S* as well as on the general modes of scientific thinking (parsimony, a questioning attitude, use of known facts, etc.) reflected in the response. To test interrater agreement, Spearman rho between scores derived from each of the two raters' ratings was used; rho was .89 for males and .62 for females.[4] Findings from this two-item test can be regarded as suggestive only; the internal consistency of the test is nil (interitem rho on the quality scores is .05 for males and −.34 for females), and findings are presented for each item separately.

*The Humor Test.* The Humor Test consisted of 15 cartoons for which *Ss* invented captions. It was administered individually with no time limit. *Ss* were asked to write out responses and to make the captions as funny as possible. Captions were rated on an eight-point scale for *humorousness*. Five raters rated the captions given by female *Ss* and these five plus two more raters rated those of male *Ss*.[5] Each rater's 15 ratings for each *S* were summed to give a final score. Interrater agreement by the intraclass correlation technique (reliability of the mean rating) was .86 for the males and .87 for the females.

*The Animal Drawing.* Each *S* was asked to draw a "fantastic animal," making it as unusual as possible. Two judges independently ranked the drawings on *drawing ability* and on *originality*.[6] Interrater reliability (rho) for drawing ability was .90 for males and .84 for females; for originality, agreement was lower (rho .29 for males and .42 for females). Disagreements were discussed until a final ranking was achieved for each variable. The final ranks for the two variables are statistically independent (rho −.01 for males and .22 for females).

*The Brick Uses Test.* The Brick Uses Test (Guilford, Frick, Christensen, & Merrifield, 1957) required *S* to list as many uses as possible for an ordinary brick in 5 minutes; scores were based on the number of shifts in the type of use given. Guilford's scoring manual permits derivation of a score having high loading on his *spontaneous flexibility* factor. Interrater agreement in the present sample was 89%.[4]

*The Consequences Test.* The Consequences Test (Guilford et al., 1957) consisted of four items, each describing some hypothetical event (for example, the sudden destruction of all books). *S*'s task was to list possible consequences of the event (time limit: 5 minutes per item). Guilford's manual gives a score having high loading on his *originality* factor. Interrater agreement here was 86%.[4]

*The Rorschach Test.* The Rorschach Test was used to derive Rorschach creativity scores as well as rankings on the independent variables: expression and control of primary process.[7]

---

[4] Both Science Test items were rated by one of the authors (FP) and one other rater. These two raters also rated the Brick Uses and Consequences Tests (to be described below). Although the author in all cases knew which *S* had produced which response, the other rater never knew this and had no way of extending a halo effect across his several ratings. Ratings were carried out by the two raters working independently and final scores were obtained through discussion and reconciliation of differences. In no case did either rater have any knowledge of the ratings of the independent variables on the Rorschach.

[5] The seven raters (none of these the authors) were all unaware of which *Ss* had produced which responses (and even of which responses were given by a single *S*). Although three of the seven had rated other creativity tests, the *Ss*' names had been removed and there was no possibility of halo effect. One of the seven raters had a partial share in rating some of the Rorschachs, but no contamination was possible since he did not know which *S* he was rating on the Humor Test.

[6] The two raters (not the authors) were each ignorant of the Rorschach scores and of which *S* had produced which drawing. These raters each contributed to the ratings of the Humor Test but to none of the other tests.

[7] Ratings of these three Rorschach variables were carried out as follows: Each protocol was scored for all variables by two raters working independently and then coming to a consensus. The two raters varied from protocol to protocol, in each case selected from a total pool of seven raters. Any problems they had in arriving at a consensus were later discussed with one of the

The creativity rating: Each Rorschach response was scored on a four-point scale. The rating is based upon a composite of statistical infrequency, sensitive use of determinants, good form level, appropriate elaboration, and a quality of richness (see Holt, 1959). To avoid contamination with the primary process scores, the *creativity* ranking in the present sample was based on the average rating of only those responses that received no primary process scores.

Amount of expression of primary process: Holt (1956; Holt & Havel, 1960) has described a theoretical rationale and an extensive scoring manual for scoring primary process manifestations in the Rorschach. All responses that have aggressive derivatives, libidinal derivatives, anxious content, or various nonlogical or bizarre formal qualities are scored. The final ranking for amount of primary process is based upon the proportion of *S*'s responses that have scorable primary process elements, double weighting for extreme instances (Level 1 as compared to Level 2 in the scoring manual). Holt and Havel (1960) report interrater agreement well into the .90's (rho) for over-all scores on amount of expression of primary process.

Effectiveness of control over primary process: Holt and Havel (1960) describe procedures for rating the "defense-demand" and the "defense-effectiveness" of each response scored for primary process. Defense-demand, rated on a six-point scale, is defined in terms of the acceptability of the response in ordinary social communication; thus, for libidinal content, for example, the response "two people holding hands" requires almost no special controls while "two people having intercourse" does. Defense-effectiveness, also rated on a six-point scale, is defined in terms of the success of the controlling factors associated with the primary process material. Such things as the form level, the associated affective response of *S*, the intellectual or esthetic context of the response, and *S*'s evasiveness or confusion are considered. Thus, the composition, "This part looks like a horse but the rest makes me think it's a man" is a less effectively controlled formal primary process response than "a centaur," which refers successfully to the social reality of a myth, puts the response at a safe distance, and makes a condensation-like response sensible. Interrater reliability of the mean defense-demand and defense-effectiveness scores was computed for the male sample; rho's were .84 and .74, respectively.

Final scores for each *S* were derived from the defense-demand and defense-effectiveness scores by a procedure developed by Goldberger (1958). For each *S*, responses where defense-demand was high and defense-effectiveness was good were tabulated; similarly, responses where defense-effectiveness was poor, whatever the defense-demand, were tabulated; responses with good defense-effectiveness but low defense-demand were omitted. Final individual scores, which were then ranked, were based on the ratio of well-controlled responses (high demand, effective defense) to poorly-controlled responses (ineffective defense, whatever the demand).

Adaptive vs. maladaptive regression: This score is based on the combined amount of expression and effectiveness of control scores described above. It was developed and used with good predictive success by Goldberger (1958), and is described in detail by Holt and Havel (1960). In brief, it produces a ranking as follows: at one extreme are *S*s who are high on both amount of expression and effectiveness of defense (controlled or adaptive regression to primary process modes of thinking); at the other extreme are *S*s who are high on amount of expression but low on effectiveness of defense (uncontrolled or maladaptive regression to primary process modes of thinking); in the middle are *S*s who are low on amount of expression (constricted control).

## RESULTS

Before studying the hypotheses of the research, we examined some of the characteristics of the rankings of Rorschach primary process. Rank correlations between the measures of expression and control of primary process were computed. The rho's were negligible (.01 for the males and .24 for the females). These correlations indicate the statistical independence of the two variables as measured here; below, we treat them as independent in correlating each separately with the quality scores. The low correlations cannot, however, be taken to reflect functional independence in the individual. Rather, we would expect that expression and control of primary process would generally be integrated in some way in an individual's functioning—but in an idiosyncratic way. It is just these idiosyncratic possibilities in the relation between the expression and control of primary process that permit statistical independence in measurement.

The combined (adaptive regression) Rorschach score correlates far more strongly with

---

authors (RRH) in a meeting of all seven raters and a final decision reached. Although the Rorschach creativity scores were derived from only those responses where primary process was not scored, it is apparent that one cannot argue that these were uninfluenced by primary process scoring in the same protocol. The Rorschach creativity score turns out to be one of the poorest correlates of the independent Rorschach variables for handling of primary process, however, and makes no significant contribution to the results (see Table 1). The two primary process ratings (expression and control) were not independently rated although they turn out to be statistically independent (see below). Of the pool of seven raters, five (including RRH) did no other ratings for the study. One was one of the seven raters for the Humor Test (where he was "blind" to *S*) and one was one of the five TAT raters (and, as described in Footnote 3, omitting her ratings made no difference in the results).

the control score component (rho is .90 for males and .79 for females) than with the expression score component (rho is .11 for males and .13 for females). Thus, on methodological grounds alone, one would expect findings for the control and the combined scores to be more similar to each other than either is to findings with the score for amount of expression. We cannot, therefore, make independent tests of Hypotheses 2 and 3 but can only look to see the directions of change (if any) in the findings for the control score when the score for amount of expression is combined with it.

Hypotheses 1, 2, and 3 state that the quality of a person's created productions is unrelated to his tendency to express more or less primary process material on the Rorschach but is related to his tendency towards more or less effective control over expressed primary process content and to a combined score (adaptive vs. maladaptive regression) taking both amount of expression and effectiveness of control into account. Thus, it was expected that the quality scores on the various experimental tests would be correlated with the control and with the combined scores but not with the score for amount of expression alone. Results are presented in Table 1.

As indicated in Row 1 of Table 1, an over-all creativity score derived for each *S* correlated significantly with the control scores for both males and females and with the adaptive regression score for males; the correlations with the amount scores, though positive, are negligible. The over-all scores were obtained by summing the ranks for each *S* on eight of the nine individual test measures and reranking the totals; to avoid double weighting the animal drawing, and since the ratings of originality had poor interrater reliability on this test, only the ranking of drawing ability was used in computing the over-all scores. In view of the variations in the subtest correlations with the Rorschach scores for the female sample (to be described below), there is some question as to the representativeness of the over-all scores for the female sample. The validity of these scores was therefore investigated.

Staff members used qualitative data from tests and life histories to *Q* sort all *S*s on

TABLE 1

RANK CORRELATIONS BETWEEN TEST MEASURES OF CREATIVITY AND THE RORSCHACH SCORES FOR PRIMARY PROCESS

| Creativity Score | Males ($N$ = 13) | | | Females ($N$ = 14) | | |
|---|---|---|---|---|---|---|
| | Amount | Control | Adaptive Regression | Amount | Control | Adaptive Regression |
| Total Creativity: over-all score | .13 | .80** | .90** | .21 | .52* | .28 |
| TAT: literary quality | .05 | .74** | .83** | .59* | .27 | .26 |
| Science (neutral): quality of theory | −.17 | .19 | .08 | −.32 | .34 | .24 |
| Science (drive): quality of theory | .11 | .43 | .40 | .64* | .11 | .01 |
| Rorschach: creativity | .17 | .27 | .31 | −.22 | .21 | .22 |
| Humor Test: humorousness | −.03 | .13 | .19 | .32 | .10 | −.13 |
| Animal Drawing: originality | −.14 | .55* | .49* | .22 | .38 | −.05 |
| Animal Drawing: drawing ability | −.04 | .69** | .68** | .24 | .39 | .33 |
| Brick Uses: flexibility | .24 | .36 | .48* | −.02 | .44 | .18 |
| Consequences: originality | −.04 | .59* | .74** | −.16 | −.02 | −.03 |

* Contrary to hypothesis. Significant at 5% level; two-tailed test.
* Significant at 5% level; one-tailed test.
** Significant at 1% level; one-tailed test.

various personality variables, which were then subjected to a syndrome analysis (Horn, 1944); syndrome scores were then converted to normative ranks (cf. Block, 1957). These rankings are partially contaminated since two of the eight component tests making up the over-all score (the TAT and the Rorschach) were used in evaluating *S*s for the *Q* sorts. With recognition of this qualification, three rankings were selected on a priori grounds to check the validity of the over-all scores. For the females, the over-all creativity score correlated with "creativity: needs for achievement, creative construction, love of novelty, sentience" (.63), with "clear and effective communication vs. blocked and uncommunicative" (.73), and with "vivid imagery vs. minimal and stereotyped fantasy and clings to reality" (.56); the corresponding correlations for the males were .61, .80, and .58, respectively. All these correlations are significant at the 5% level or better (two-tailed test). Thus, the over-all scores have some measure of validity. Be this as it may, valid test differences may still be obscured by these over-all scores.

Correlations between the Rorschach variables on the one hand and the quality scores on the individual tests on the other are presented in the remainder of Table 1. For the males, the correlations test by test parallel those with the over-all score. In each of the nine measures, the quality score correlates more highly with the Rorschach control score than with the amount score. In four instances, the correlation with control is significant and in all cases it is positive; in five instances, the correlation with amount is negative and in all cases it is negligible. The score for adaptive vs. maladaptive regression tends to give slightly higher correlations than the control score; the correlations are significant in five instances. There is little patterning among the tests that do and do not produce significant correlations; tests as different as the TAT, the animal drawing, and the Consequences Test give the highest correlations.

For the females, there is much more variation in the separate test correlations. Only two correlations are significant, and these—surprisingly—are both with the amount of primary process expression. Correlations with control and with the adaptive regression score are higher than those with amount in only about half of the cases.

Since other data available on this same female sample suggest that the adaptive regression score is not a meaningful one for this particular group (see Holt & Havel, 1960), and since this score produced no significant correlations in the present sample, we focus then on the two simpler (and independent) Rorschach scores, amount and control, for the females. The major difference in results between males and females is in the correlations with the amount score. For both groups, the correlations with control tend to be consistently positive (except for the Consequences Test for females) although the correlations for the females are smaller; that the females had less variation in their control scores may account for some of this difference. The wide range in correlations with amount in the female group, from strong positive to moderate negative, sharply differentiates the two groups.

The correlations nevertheless seem to fall into a pattern for the females. The tests which, at least in appearance, permit more open-ended and expressive verbal responses, and in particular permit more drive expression in the responses, are just those where the correlation for the female *S*s is higher with the Rorschach score for amount of primary process expression than with the control score. This category includes the TAT and the drive related Science Test item (and perhaps the Humor Test).[8] Why should it be that, for our college girls, individual differences in amount of primary process expression are at times better predictors of the quality of imaginative productions, and individual differences in control over expressed primary process material are at other times better predictors, while in the males the individual differences in control over primary process expression are consistently the better predictors of the quality scores? A critical point here, we

[8] On the face of it, the Rorschach test is also such an open-ended test and permits drive expression. It was not used in this analysis, however, because the Rorschach creativity score is based on only those responses where primary process is not scorable in the response and because a control factor (form level) is built into the rating criteria.

suggest, is that people vary in their automatic (not necessarily thought out) approach to a task in terms of the implicit assumptions they hold regarding what the task and the experimenter requires of them. Thus, for tests taken in an experimental situation and in a university setting, *S*s may implicitly "direct" (or "set") themselves to give highly controlled imaginative productions; in this case, individual differences in the capacity to exert effective control over imaginative productions would get reflected in the quality of the finished product. In contrast, given such tests as the TAT, and in spite of the university surroundings, *S*s may at times approach test tasks as though the appropriate response were to give a fanciful and freely expressive response, even if the response would not be too well controlled; in this case, individual differences in the capacity for fanciful expression would be reflected in the quality of the finished product. In our data, it is as though the males consistently had the former (control) response set while the females had at times the former and at times the latter response set.

Why these college girls might have demonstrated such duality in their implicit approach to the tasks is of interest, and may reflect a greater field dependence (since they vary with the ostensible task "demand"), or a bimodal identity including both more expressive and more controlled aspects (feminine-masculine) with neither being consistently more dominant. Be this as it may, these speculations led to some expectations regarding the data. If the males responded to even the more fanciful tests in a way reflecting the effectiveness of their controls, and if the females responded to these tests in a way reflecting their potential for fanciful expression, then there should be two ways to get high (or low) quality scores on these tests: either by effectively (or ineffectively) controlled imaginative productions or by highly (or minimally) expressive productions. The males should tend to give the former, and the females the latter; but this should not be the case for the tests which "demand" more specific and controlled productions. Some indirectly relevant data were available for two of the more expressive tests (TAT and drive related Science) and for one of the more controlled ones (neutral Science). *S*s had been ranked for the *amount* of drive content in their TAT and Science productions and also for the degree of *control* over expressed drive content (see Pine, 1959, 1960); control was reflected principally in an integration of the drive content into the core of the created product. These ratings of expression and control of drive content are not identical with the ratings of expression and control of primary process on the Rorschach because of test differences and because the Rorschach ratings included formal as well as drive indices of primary process (although the latter predominated); however, what is more important, they do reflect expression and control within these created productions themselves and thus provide a better test of the hypothesis that there are two kinds of productions which could get high scores. It was expected that the quality scores on the TAT and the drive related Science item would correlate more highly with control of drive for the males than for the females and more highly with amount of drive expression for the females than for the males; this difference was not expected to appear in the neutral Science item. Correlations (rho) are presented in Table 2.

Correlations with amount of drive are higher for the females than for the males on the TAT and the drive related Science item but not on the neutral Science item. Correlations with the control scores are higher for the males than for the females on the TAT but the reverse is true on the neutral Science item. (Throughout, larger group differences appear in the Science items than in the TAT.) Thus, there is some indirect support for the proposition that the females achieved high (or low) quality scores on the more open-ended, expressive tests in a somewhat different way from

TABLE 2

RANK CORRELATIONS BETWEEN THE QUALITY SCORES AND EXPRESSION AND CONTROL OF DRIVE CONTENT IN THE TAT AND SCIENCE TEST

| | TAT | | Science (drive) | | Science (neutral) | |
|---|---|---|---|---|---|---|
| | Amount | Control | Amount | Control | Amount | Control |
| Males | .60 | .42 | .09 | .05 | .12 | .06 |
| Females | .77 | .28 | .52 | .04 | −.36 | .64 |

the males; individual differences in drive expression were more significant in the females than in the males on these tests.

The correlations shown in Table 2 are also of interest for their parallel to the correlations of Table 1. In both, for the male sample, the TAT produces a significant correlation with the control score (although the correlation with amount is higher in Table 2), while the two Science items produce no significant correlations. The variation in the correlations for the females found in Table 1 is repeated almost exactly in Table 2. In the TAT and the drive related Science item, all correlations are positive and the correlation of the quality score with amount is much higher than with control; in the neutral Science item, the correlation with control is positive and with amount is negative. Thus, with amount and control scores derived from the TAT and the Science productions themselves as well as with the parallel scores on the Rorschach test the same sex differences and patterning of correlations are obtained. These findings support the validity of these patterns of difference in the present sample.

## DISCUSSION

The results are suggestive in several ways. First, the way in which primary process material is handled on the Rorschach test is related to the quality of imaginative production in a variety of test situations. That control over primary process expression, rather than gross amount of expression per se, is related to the quality of created products is indicated by all of the data of the male sample and by results with the over-all creativity score for the females. These results give general support to the hypotheses of the study, and they indicate further that Kris' concept of regression in the service of the ego can be subjected to empirical study—that predictions generated by this psychoanalytic concept can be studied with operationally defined scores and quantitative techniques.

The data for our male sample, where scores for adaptive regression and for controlled primary process expression could predict quality scores on tests of imagination even where primary process thinking was not evidently required by the tests, are of additional interest. The tests used here have certain parallels to the model of primary process thinking: thus, the TAT requires a "make-believe," nonliteral kind of thinking and also the use of drive content (for motivation of characters); the role of primary process in humor has been discussed by Freud (1938); even with the Brick Uses there is a parallel to primary process in that high scores are attained by *shifts* in the direction of thinking, by a style that contrasts with the directed thought of the secondary process. But these parallels between the test requirements and primary process thinking are at best tenuous in most or all of the tests used. We would argue, in contrast, that primary process thinking is not directly implicated in the production of responses to these tests, but rather that modes of expression and control of primary process (perhaps beginning in early childhood) become generalized as broad cognitive styles (cf. Klein, 1950) which are reflected in all areas of thinking. Thus, where tests require flexible and original thinking, and where the control style permits this, thinking can be flexible even apart from the (hypothesized) original relation of control directly to primary process thinking.

But there is an important caution to be drawn from our results. The subsample differences indicate that there are limitations to the extent to which we can generalize from these data. Kris' discussion of the relation of regression in the service of the ego to creativity can be viewed as a broad approximation—an approximation subject to important differences in samples that vary as to age, sex, the absolute level of creativity of the *S*s involved, and the nature of the creative production. The most we can say from our data is that expression and control of primary process are relevant to an understanding of the creative process, but that the specific nature of the relationships involved may vary and awaits much further research for detailed specification.

The sex differences in our data are of interest for their implications regarding the relationship between "creativity" on the present test measures and creativity as it is more naturally found in the productive arts and sciences.

First: Do the two significant correlations between quality scores and *amount* of primary process expression (rather than with effective-

ness of control) indicate that some highly rated productions were quite expressive but seriously lacking in controls? This would contrast to an "ideal" of creative productivity as both personally expressive and sufficiently controlled to be communicable and potentially socially valued. Whether or not some truly creative productions can be really uncontrolled, or whether the nature of controls in some areas (e.g., painting, dance) is simply different from and perhaps more subtle than the intellectual verbal controls required by most of our tests, the essential point is that our data do not really suggest that some high quality productions here were uncontrolled. The *S*s who gave these results were "normal" college girls. It can be assumed that some control factors were included in *all* of their expressions of primary process, and the Rorschach scores support this assumption. Furthermore, these girls are obviously all sufficiently well controlled to function effectively in school and generally, and presumably they would be able to impose control when the adaptive requirements of a task require this (as in the tests of imagination) even if some of them could relinquish control somewhat on the Rorschach.

Second: The introduction of the idea that our male and female *S*s had implicit but differing response sets which led them to approach the various tasks differentially requires comment. Certainly this concept is relevant for our present sample where sudden "inspiration" had little or no part in the production (upon demand) of imaginative products and where varying expectations and goals were likely to be activated by the entire experimental setting. But does such variation in approach to productive work, variation in terms of more expressive or more controlled responsiveness, apply in the creative arts and sciences? The concept certainly does not seem applicable where creative experiences come in off-guard moments, during sleep, or when the sudden inspiration is experienced as though coming upon the person from outside without his being responsible for it (cf. Ghiselin, 1952). But there is a way in which individual preferences for more expressive or more controlled productivity are relevant for the broader realms of creativity: in the way in which people, consciously or not, gravitate towards areas conducive to differing kinds of productivity. Thus, as the girls in the present sample may have responded with more or less fanciful and idiosyncratic expression in tasks which permitted such expression, so there seem to be differences in western culture in the areas where males and females have made their greatest contributions. A Sappho existed long before a Madame Curie, and it is probably still true that there is a greater proportion of women in the arts than in the sciences. Thus, alternative areas open for productive work in a particular social milieu may provide room for differing individual styles of creativeness.

## SUMMARY

The present paper reports a study of relationships between expression and control of primary process material on the one hand and quality of productions in a variety of tests of imagination on the other. Scores reflecting the quality of imaginative productions were derived from the following tests: TAT, Science Test, Humor Test, Rorschach, Animal Drawing, and Guilford's Brick Uses and Consequences Tests. Holt's manual was used in assessing expression and control of primary process in the Rorschach test. *S*s were 13 male and 14 female undergraduates.

Amount of expression of primary process and effectiveness of control over such expression were shown to be statistically independent. It was expected that the quality scores on the tests of imagination would be unrelated to amount of primary process expression and positively related to the effectiveness with which such expression is controlled. Results on all tests for the males, as well as over-all scores for both groups, supported the hypothesis. The correlations for the individual tests for the females, however, showed many unexpected variations.

An hypothesis regarding the results with the female sample was advanced and explored—based on sex differences in implicit orientation to the tasks which produced relative differences in the expressiveness and control of responses in males and females. Implications of the findings in relation to creativity as it is encountered in natural, nonexperimental, settings were discussed.

## REFERENCES


BLOCK, J. A comparison of ipsative and normative ratings of personality. *J. abnorm. soc. Psychol.*, 1957, **54**, 50–54.

EBEL, R. L. Estimation of the reliability of ratings. *Psychometrika*, 1951, **16**, 407–424.

FREUD, S. *Delusion and dream*. London: Allen & Unwin, 1921.

FREUD, S. Wit and its relation to the unconscious. (German ed., 1905) In A. A. Brill (Ed.), *Basic writings of Sigmund Freud*. New York: Modern Library, 1938.

FREUD, S. Formulations regarding the two principles in mental functioning. (German ed., 1911) In, *Collected Papers*. Vol. IV. London: Hogarth Press, 1953. Pp. 13–21.

FREUD, S. *The interpretation of dreams*. (German ed., 1900) New York: Basic Books, 1955.

GHISELIN, B. *The creative process*. Los Angeles: Univer. California Press, 1952.

GOLDBERGER, L. Individual differences in the effects of perceptual isolation as related to Rorschach manifestations of the primary process. Unpublished doctoral dissertaion, New York University, 1958.

GOLDBERGER, L., & HOLT, R. R. Experimental interference with reality contact: Individual differences. In P. Solomon (Ed.), *Sensory deprivation*. Cambridge: Harvard Univer. Press, 1961, in press.

GUILFORD, J. P., FRICK, J. W., CHRISTENSEN, P. R., & MERRIFIELD, P. R. A factor-analytic study of flexibility in thinking. *U. Stn. Calif. Psychol. Lab. Rep.*, 1957, No. 18.

HOLT, R. R. Gauging primary and secondary processes in Rorschach responses. *J. proj. Tech.*, 1956, **20**, 14–25.

HOLT, R. R. Manual for the scoring of primary process manifestations in Rorschach responses. (Draft 7) New York: Research Center for Mental Health, 1959. (Dittoed)

HOLT, R. R., & HAVEL, JOAN. A method for assessing primary and secondary process in the Rorschach. In Maria A. Rickers-Ovsiankina (Ed.), *Rorschach psychology*. New York: Wiley, 1960. Pp. 263–315.

HORN, D. A study of personality syndromes. *Character & Pers.*, 1944, **12**, 257–272.

KLEIN, G. S. The personal world through perception. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: An approach to personality*. New York: Ronald, 1950. Pp. 328–355.

KRIS, E. *Psychoanalytic explorations in art*. New York: International Univer. Press, 1952.

MILNER, MARION. *On not being able to paint*. New York: International Univer. Press, 1957.

PAPPENHEIM, ELSE, & KRIS, E. The function of drawings and the meaning of the "creative spell" in a schizophrenic artist. *Psychoanal. Quart.*, 1946, **15**, 6–31.

PINE, F. Thematic drive content and creativity. *J. Pers.*, 1959, **27**, 136–151.

PINE, F. A manual for rating drive content in the Thematic Apperception Test. *J. proj. Tech.*, 1960, **24**, 32–45.

RAPAPORT, D. Toward a theory of thinking. In D. Rapaport (Ed.), *Organization and pathology of thought*. New York: Columbia Univer. Press, 1951. Pp. 689–730.

SCHAFER, R. Regression in the service of the ego: The relevance of a psychoanalytic concept for personality assessment. In G. Lindzey (Ed.), *Assessment of human motives*. New York: Rinehart, 1958. Pp. 119–148.


(Received August 17, 1959)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 380-388

# EFFECTS OF STRESS ON GOAL SETTING BEHAVIOR

KUNAPULI U. RAO AND ROGER W. RUSSELL[1]
*Government Training College, Adnhra Pradesh, India* *Indiana University*

In an earlier paper one of us (Russell, 1953) discussed analogies between the concept of "stress" as it has been applied to the dynamics of physical, biological, and behavioral systems. Many of these are systems that tend to maintain relatively steady states over extended periods of time. External forces, stresses, impinging upon the system are balanced by changes, strains, within the system. More recently Davis (1958) has described this dynamic process in terms of "negative feedback": "[The] basic requirement [of negative feedback] is that some energy is taken from a later part of the system and introduced into an earlier part in such a way as to oppose the change produced there by the original input" (p. 10). The relation between input and output, or between stress and strain, holds within limits; beyond the limits, however, stresses may produce temporary or permanent distortions of the system.

Applying this model to behavior systems suggests that disorders of behavior may occur when responses are inadequate to compensate for changes in the systems produced by stresses impinging upon them. On a priori grounds it might be predicted that this state of affairs would exist when the organism does not possess the capabilities or skills required for an adequate response or when competition arises between two or more incompatible response tendencies, resulting in delay or failure in the occurrence of an adequate response. In the psychological literature the former condition has frequently been referred to by the generic term, "frustration," and the latter by "conflict." Experimental studies have shown that exposure of a variety of living organisms to either of the conditions may indeed produce behavior disorders (Russell, 1950); however, distortions of behavior systems are by no means inevitable results in such experiments. Distortion would not be expected to occur if feedback from the performance or output of a system altered the system in such a direction that its response capabilities were adequate to compensate for changes produced by future inputs of a similar nature. The present experiment was designed to study a familiar and frequently occurring situation in which behavior systems are altered in this way.

It is a commonly observed characteristic of human behavior that a person when frustrated by failure to achieve a goal he has set for himself tends to alter the goal in a direction more consistent with his actual performance. This is a process in which feedback, knowledge of results, from the output, performance, produces changes in the behavior system, i.e., changes in goals; the latter, in turn, affect the future response of the system. We have seen this goal setting process in action in experimental situations designed to impose stress upon our subjects; its effect has been to reduce or to eliminate predicted disorganization of behavior. In the light of clinical as well as experimental evidence it appears that goal setting is an important process in preventing disorganization of behavior and, therefore, deserves special attention in experimental psychopathology.

The process of goal setting may be studied experimentally using the familiar level of aspiration technique (see Frank, 1935b; Lewin, Dembo, Festinger, & Sears, 1944; Rotter, 1954). The technique has three major features: at the beginning of each trial the subject states the goal he expects to achieve during the trial; he then performs the required task; and, finally, he obtains knowledge of the results of his performance. Knowledge of results provides a feedback the effects of which on stated levels of aspiration and performance may be observed in subsequent trials. In the present experiment we have used this behavior system to study the effects on goal setting behavior of stress imposed on the

[1] The research was conducted at the Department of Psychology, University College, London, England. We wish to express our appreciation to A. R. Jonckheere and A. Summerfield for their critical examination of our methods for analyzing results of the study and to the Headmaster of the school whose very kind cooperation facilitated the collection of our data.

system by failure to achieve stated goals. By controlling performance without the subject's knowledge it has been possible to vary systematically the number of successive trials during which failure occurred. We have also studied the aftereffects of the stress on levels of aspiration and on performance during a recovery period when failure was not induced experimentally.

The model in terms of which we have viewed the goal setting process permits certain predictions that should be testable under the conditions of the present experiment:

1. The goal setting process should lead to stated levels of aspiration that are consistent with performance and remain relatively stable as long as performance is stable. This is, of course, the familiar result of many studies employing the level of aspiration technique.

2. Stressing the behavior system by introducing repeated failure to achieve stated goals should upset the stability of the prestress levels of aspiration and lead toward the establishment of new levels. This prediction is also supported by evidence from a number of previous experiments in which it has been a consistent finding that level of aspiration tends to be raised after successful performance and lowered after failure.

3. There should be an inverse relation between duration of exposure to stress and level of aspiration such that the level decreases as the number of stress trials increases. We are particularly interested in the nature of this relation and our research was designed to provide an empirical description of it.

4. Return to prestress experimental conditions should be reflected in shifts of stated levels of aspiration in the direction of their original levels of stability. Here we are particularly interested in two characteristics of the recovery process: (*a*) how is rate of recovery related to the duration of previous exposure to stress; and (*b*) is there a tendency toward "overshooting" and oscillation such as may occur in other feedback systems (Davis, 1958; Selye, 1946)?

## METHOD

### *Experimental Design*

The general plan of the research was to obtain measures of level of aspiration and of performance before and after exposure to varying numbers of trials in which subjects consistently failed to achieve their stated goals. The plan also included measurement of level of aspiration during the stress trials; performance could not be measured during this phase of the research since the experimenter manipulated performance scores artificially in order to produce consistent failure to achieve levels of aspiration.

The experimental design included four phases, always presented in the same temporal order. It required four groups of experimental and four of control subjects, all groups matched initially on variables likely to affect behavior in the research situation. In Phase 3, the four experimental groups were subjected to "stress trials" totaling 10, 20, 30, and 40 trials, respectively, while the control groups performed corresponding numbers of "standard trials." In all other phases experimental and control groups received identical treatments.

### *Subjects*

The subjects were 160 adolescent boys, chosen from a secondary technical school. The experiment was planned in two replications, the first consisting of 96 and the second of 64 subjects. The former were pupils of the third form during the school year 1952–53; the latter, pupils of the same form admitted during 1953–54.

Previous research (see Lewin et al., 1944; Rotter, 1942a) has shown that age, sex, socioeconomic background, past success and failure, and intelligence are major variables affecting goal setting behavior in the level of aspiration situation. Therefore, it was necessary to match the eight groups required by the experimental design on these variables. Possible effects of sex differences were controlled by using only male subjects. All boys selected as subjects belonged to families of the socioeconomic class, "manual workers" (Mosier & Hall, 1954). To control for possible effects of past academic success or failure only boys with satisfactory academic progress during the preceding school year participated in the experiment. Table 1 shows that there were no significant differences among the groups in their chronological ages, intelligence quotients,[3] nor in their initial scores on measures of the behavior which constituted the dependent variables in the experiment.

### *Apparatus and Instructions*

The apparatus consisted of a slightly modified form of the Rotter Aspiration Board (Rotter, 1942b). It was used to present a simple motor task in which the subject was required to use a small cue to hit a steel ball along a groove cut into a wooden base block. Runners attached to both sides of the groove kept the ball in the channel. Marked along the two sides of the channel were 19 units, numbered in increasing order from 1 to 10 and then decreasing to 1. In the channel at points opposite the center of each unit was a circular indentation, in one of which the ball came to rest after each hit.

[3] Intelligence quotients were derived from scores on the Otis Self-Administering Test, Intermediate Examination, Form A, as adapted by the Australian Council for Educational Research.

TABLE 1
MATCHING OF CONTROL AND EXPERIMENTAL GROUPS

| Variable | $F$[a] | $df$ | Mean |
|---|---|---|---|
| Chronological age (months) | 0.57 | 7 and 152 | 166.6 |
| Intelligence quotient | 1.58 | 7 and 152 | 106.6 |
| Initial level of aspiration | 1.59 | 7 and 152 | 30.3 |
| Initial performance | 0.64 | 7 and 152 | 27.7 |

[a] None of these $F$ ratios is significant.

In Phase 3 of the experiment it was necessary for the experimenter to control the point where the ball came to rest. This was accomplished by altering the relative positions of the runners along the sides of the channel, thereby narrowing the channel and jamming the ball. The apparatus was so designed that this could be done without the knowledge of the subject. One runner consisted of two pieces, one piece pivoted at its central end and the other end joined to a Bowden cable by means of a bracket. The second runner was also pivoted at one end and joined to the cable at the other. By turning the cable control handle to different extents the ball could be stopped at various points along the channel.

In accordance with the practice adopted by previous investigators using this task, five hits were taken to constitute one trial. The sum of the scores for the five hits provides a score that is more stable than scores for individual hits. Before each hit the steel ball was placed at a standard point at the subject's end of the channel. Each trial of five hits began with the subject's stated level of aspiration, i.e., the number of points he expected to obtain during the trial. The subject then hit the ball with the cue. After each hit, the subject and the experimenter could determine the score in terms of the number of the unit on the channel in which the ball came to rest.

Rotter (1942a, 1942b) has shown that the instructions given subjects regarding the task have important influences upon the results obtained. Therefore, the standardized instructions used in the present study conformed to those suggested by his research. Precautions were taken to be certain that the instructions were clearly understood. After they were given, the subject was provided an opportunity to practice. This practice consisted of 10 hits for each; it allowed the experimenter to observe whether or not the subject could apply the instructions correctly. Immediately after this practice period the main features of the instructions were repeated and a standard series of questions asked, the answers to which indicated whether or not the subject had understood. Further explanation was given if necessary.

### *Procedure*

As far as possible subjects from the various groups were taken in a random order, but sometimes this order could not be adhered to because of absences, school examinations, etc. The experiment was conducted at the school in a private room made available for the purpose.

*Phase 1: Practice.* When a subject arrived, he was seated in a chair at the subject's end of the apparatus and given the general instructions, the 10 practice hits, and a review of the instructions as described above.

Three main steps were taken to insure that the subjects were as highly motivated as possible during the experimental period.

1. The experiment was introduced as a new game.
2. Subjects were told that their scores were measures of their general ability and would be seen by the Headmaster of the school.
3. Monetary prizes were offered and given when the study was completed, the amounts being set after consultation with the Headmaster and taking into consideration the extent of funds normally available to schoolboys as "spending money." Replies to a standard series of questions, put to each subject when his experimental session had been completed, suggested that all worked for good scores. They appeared to be more concerned with performing well in order to gain approval of the experimenter and the Headmaster than to obtain the monetary rewards.

*Phase 2: Prestress.* During Phase 2, the prestress phase, all subjects were given 15 standard trials of five hits each. The procedure for each trial contained three basic steps. The subject first stated his level of aspiration, the score he expected to obtain, for that trial; he then performed the task; and finally he was given knowledge of his performance scores.

*Phase 3: Stress.* Two independent variables, presence or absence of stress and number of stress trials, were introduced in this phase of the experiment. As stated earlier, stress was defined as trials during which the subject consistently failed to achieve his stated levels of aspiration. Performances during stress trials could be controlled since the experimenter manipulated the apparatus without the subject's knowledge, thus determining at which unit in the channel the ball would come to rest after each hit. All experimental groups were exposed to such stress, successive groups receiving 10, 20, 30, and 40 stress trials, respectively. The various control groups continued the standard procedure, as in Phase 2, for numbers of trials corresponding to their paired experimental groups.

*Phase 4: Recovery.* During the final phase of the experiment all groups were given 40 standard trials, following the procedure used in Phase 2.

Before leaving the experimental room each subject was asked not to discuss the experiment with anyone in the school and, as a token of his willingness to conform to this request, signed his name on the experimenter's score sheet. Examination of performance scores and of replies to questions at the end of the experimental session strongly suggest that the subjects kept strictly to their promises and that no information about the experiment leaked out. The possibility of losing one of the larger monetary rewards to another boy with prior information may also have helped in achieving this end.

## RESULTS

Results obtained in the various phases of the experiment are described in the following paragraphs. Since our principal interest in the study was in the goal setting process, the data

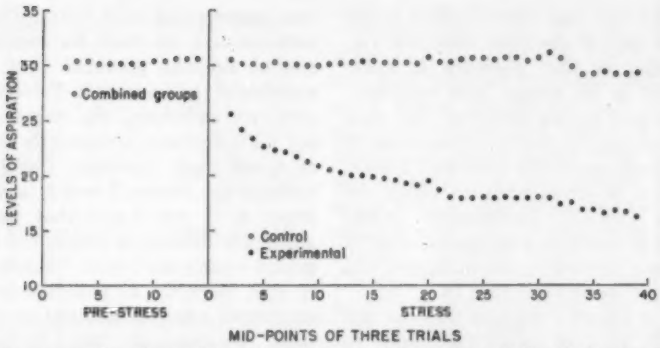


FIG. 1. Curves representing three-point moving averages of levels of aspiration during prestress and stress trials.

for level of aspiration and performance are presented separately. Throughout the report, "significance" is defined conventionally in terms of the .05 level of confidence.

*Prestress*

*Goal setting behavior.* A general picture of goal setting behavior when the subjects worked freely at the experimental task is shown by the curve for combined groups on the left side of Figure 1. This, as all other curves presented in the report, is expressed in terms of three-point moving averages in order to show clearly the trends of the data. Prior to the prestress phase of the study, all subjects received 10 practice hits; this practice was sufficient to have resulted in a level of aspiration that remained extremely stable for all experimental and control groups during the 15 prestress trials. Table 2 presents the mean prestress levels of aspiration for all groups; none of the very small intergroup differences was significant.

*Performance.* Figure 2 shows that performance also reached a stable state very early in the prestress trials. There were no significant intergroup differences in the mean performance scores, which are presented in Table 3.

*Stress*

*Goal setting behavior.* Levels of aspiration for all control groups continued to remain stable throughout the next phase of the study, when experimental groups were being subjected to stress. Each of the four control groups received different numbers of trials, i.e., 10, 20, 30, and 40, during this phase, but, in order to show a composite picture of their goal setting behavior, the scores have been combined in constructing the curve shown in Figure 1. There were no changes in the conditions under which the control subjects operated and, as Figure 1 indicates, their levels of aspiration continued at the same steady state level as during the prestress phase.

Conditions were changed for the four experimental groups, who received 10, 20, 30, and 40 successive trials, respectively, during all of which the subjects failed to achieve their stated levels of aspiration. As Figure 1 shows, the change was reflected immediately in the subjects' goal setting behavior. Stated levels of aspiration always decreased on successive trials, although at an apparently decelerating rate as the number of stress trials increased.

There are several possible ways of expressing these descriptive trends in numerical terms. The method chosen for the present analysis provided a "net score" calculated as follows: The mean level of aspiration for each subject in the prestress phase was calculated. Differences between this standard and levels at subsequent stages in the experimental session were determined as measures of the increase or decrease in level associated with the experimental treatment. The differences between such measures for an experimental subject and his corresponding control provided "net scores." Table 4 presents the median net decreases in levels of aspiration of the four experimental groups during the stress phase of

the study. The fact that these derived scores for all groups were in the same direction, i.e., decreases, indicates that exposure to stress was associated in all groups with consistent alterations in goal setting behavior. The data in Table 4 also suggest that the magnitude of these decreases in aspiration level were related to the number of stress trials to which the groups were exposed. Jonckheere's (1954) distribution-free *k*-sample test against ordered alternatives was applied to the data, and this trend was found to be significant ($p < .01$).

The curve in Figure 1 suggests that the rate of decrease in level of aspiration tended to decelerate with increasing duration of exposure to stress. This curvilinear relation is shown in Table 5, which presents the mean net shifts in levels during successive blocks of 10 trials. The means were derived in the following way: For Block 1 the shift for each subject in the four experimental groups was obtained by finding the difference between his mean level during prestress trials and that during the first 10 stress trials. Similar calculations for corresponding control subjects provided measures representing their shifts. The differences between such measures for experimental and control subjects provided "net shifts." Inspection of the columns in Table 5 shows that, with one exception, the magnitude of these net shifts decreases consistently as the number of stress trials increases. Using analysis of variance for Blocks 1 and 2, and a *t* test for Block 3, it was found that there were no significant differences among the experimental groups in any one block. Therefore, the shifts in each block for all groups were pooled and an overall mean determined as shown in the final column of the table. Applying Jonckheere's (1954) test, the trend for the overall mean shift to decrease as the number of stress trials increased was found to be significant ($p < .01$). In other words, the effects of continued exposure to stress diminished significantly as the total number of stress trials increased.

It is, of course, impossible to present a similar analysis of performance scores during the stress phase of the study since the experimental design required that all trials for the experimental groups be influenced artificially by the experimenter in order to produce consistent failure. Without meaningful performance scores it is also impossible to present the derived "difference" or "goal discrepancy" scores (Frank, 1935a; Lewin et al., 1944) frequently reported by investigators to express the relation between a level of aspiration and the level of performance that preceded it. Since we are primarily interested in the goal setting process itself, such derived scores are not essential to our report.

TABLE 2
MEAN LEVELS OF ASPIRATION: PRESTRESS PHASE

| Replication | Groups | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $C_1$ | $E_1$ | $C_2$ | $E_2$ | $C_3$ | $E_3$ | $C_4$ | $E_4$ |
| 1 | 30.1 | 30.8 | 29.4 | 30.6 | 31.2 | 31.6 | 29.8 | 31.0 |
| 2 | 30.7 | 31.2 | 29.4 | 29.6 | 30.8 | 29.3 | 28.3 | 29.8 |
| Combined[a] | 30.3 | 30.9 | 29.4 | 30.2 | 31.1 | 30.6 | 29.2 | 30.5 |

[a] The differences among groups are not significant ($F = 1.59$, $df = 7$ and 152).

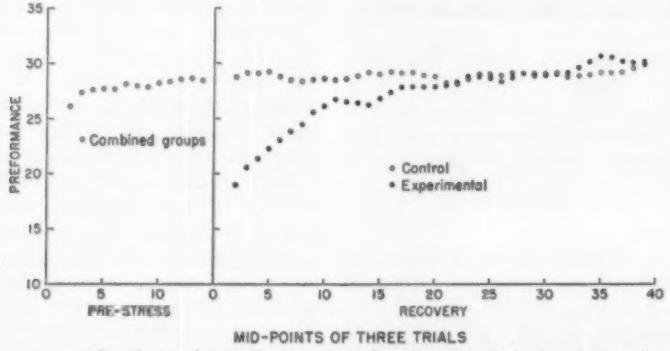


FIG. 2. Curves representing three-point moving averages of performance during prestress and recovery trials.

TABLE 3
MEAN PERFORMANCE SCORES: PRESTRESS PHASE

| Replication | Groups | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $C_1$ | $E_1$ | $C_2$ | $E_2$ | $C_3$ | $E_3$ | $C_4$ | $E_4$ |
| 1 | 27.7 | 26.8 | 27.8 | 27.4 | 27.4 | 28.2 | 27.4 | 28.0 |
| 2 | 27.6 | 28.3 | 27.8 | 26.9 | 28.6 | 28.0 | 27.1 | 27.8 |
| Combined[a] | 27.6 | 27.4 | 27.8 | 27.2 | 27.9 | 28.1 | 27.3 | 27.9 |

[a] The differences among groups are not significant ($F = 0.64$, $df = 7$ and 152).

*Recovery*

*Goal setting behavior.* At the start of the fourth phase of the study changes in operating conditions were introduced for the experimental groups by allowing the subjects once again to work freely at the experimental task; control subjects still operated under the standard conditions that had prevailed for them throughout the experiment. Figure 3 shows that the levels of aspiration of the control subjects continued at the same steady state level as during the preceding two phases of the study; for the four control groups the steady state had persisted for 65, 75, 85, and 95 trials, respectively.

On the other hand, removal of the stress conditions introduced a rapid change in the stated levels of aspiration for all the experimental groups. The ascending curve in Figure 3 shows the general trends of the goal setting process. The process resulted in recovery to prestress levels of aspiration, levels which had characterized the subjects' behavior under conditions identical with those now existing and which corresponded to those maintained by the control groups throughout the experiment. The sensitivity of the goal setting process to the change in operating conditions is shown by the rapid shift from a three-point moving average of 16.1 at the end of the stress trials to an average of 20.8 at the beginning of the recovery trials.

Figure 3 is a composite curve constructed by combining scores for subjects in all four experimental groups. Although it shows general trends, it does disguise one kind of information that is clearly apparent when curves are plotted for each experimental group individually, i.e., the duration of the recovery process appears to be related to the number of stress trials imposed during the preceding phase of the experiment. This relation can be expressed numerically in several ways, all of which give similar results. The data we report here were obtained by a simple graphic method using curves plotted for each experimental group individually: the number of the trial on the abscissa corresponding to the first point on the recovery curve that equaled the prestress mean level of aspiration for the group was taken as representing the number of trials for recovery. The data are presented in Table 6. The trend toward increasing number of trials to recover with increasing number of stress trials which is apparent in these data was found to be significant when tested by Jonckheere's (1954) method and by analysis of variance: by both methods $p < .01$. A more detailed examination of the trend suggests that it is not rectilinear, but tends to approach an asymptote between 30 and 40 stress trials. This possibility was examined by $t$ tests between all the various pairs of experimental groups, using data for the combined replications. The results of the tests are given in Table 7 and show that, as inspection of the basic data had suggested, all differences in number of trials to recover were significant except the difference between Groups $E_3$ and $E_4$, the groups with 30 and 40 stress trials, respectively.

We wish to call attention to one other feature of the recovery curve for the experimental groups. Examination of the curve in Figure 3 suggests that there was some oscillation and "overshooting" in the recovery process after the prestress levels of aspiration had been attained. Had we had some earlier indication of when in the recovery trials these phenomena might have occurred, we would have extended the recovery phase with the intention of determining if and when the phenomena disappeared and a final steady state reached.

TABLE 4
"NET DECREASES" IN LEVELS OF ASPIRATION OF EXPERIMENTAL GROUPS: STRESS PHASE

| Groups | Number of Stress Trials | Median Score |
|---|---|---|
| $E_1$ | 10 | 7.5 |
| $E_2$ | 20 | 9.2 |
| $E_3$ | 30 | 10.2 |
| $E_4$ | 40 | 11.7 |

the study. The fact that these derived scores for all groups were in the same direction, i.e., decreases, indicates that exposure to stress was associated in all groups with consistent alterations in goal setting behavior. The data in Table 4 also suggest that the magnitude of these decreases in aspiration level were related to the number of stress trials to which the groups were exposed. Jonckheere's (1954) distribution-free $k$-sample test against ordered alternatives was applied to the data, and this trend was found to be significant ($p < .01$).

The curve in Figure 1 suggests that the rate of decrease in level of aspiration tended to decelerate with increasing duration of exposure to stress. This curvilinear relation is shown in Table 5, which presents the mean net shifts in levels during successive blocks of 10 trials. The means were derived in the following way: For Block 1 the shift for each subject in the four experimental groups was obtained by finding the difference between his mean level during prestress trials and that during the first 10 stress trials. Similar calculations for corresponding control subjects provided measures representing their shifts. The differences between such measures for experimental and control subjects provided "net shifts." Inspection of the columns in Table 5 shows that, with one exception, the magnitude of these net shifts decreases consistently as the number of stress trials increases. Using analysis of variance for Blocks 1 and 2, and a $t$ test for Block 3, it was found that there were no significant differences among the experimental groups in any one block. Therefore, the shifts in each block for all groups were pooled and an overall mean determined as shown in the final column of the table. Applying Jonckheere's (1954) test, the trend for the overall mean shift to decrease as the number of stress trials increased was found to be significant ($p < .01$). In other words, the effects of continued exposure to stress diminished significantly as the total number of stress trials increased.

It is, of course, impossible to present a similar analysis of performance scores during the stress phase of the study since the experimental design required that all trials for the experimental groups be influenced artificially by the experimenter in order to produce consistent failure. Without meaningful performance scores it is also impossible to present the derived "difference" or "goal discrepancy" scores (Frank, 1935a; Lewin et al., 1944) frequently reported by investigators to express the relation between a level of aspiration and the level of performance that preceded it. Since we are primarily interested in the goal setting process itself, such derived scores are not essential to our report.

TABLE 2
MEAN LEVELS OF ASPIRATION: PRESTRESS PHASE

| Replication | Groups | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $C_1$ | $E_1$ | $C_2$ | $E_2$ | $C_3$ | $E_3$ | $C_4$ | $E_4$ |
| 1 | 30.1 | 30.8 | 29.4 | 30.6 | 31.2 | 31.6 | 29.8 | 31.0 |
| 2 | 30.7 | 31.2 | 29.4 | 29.6 | 30.8 | 29.3 | 28.3 | 29.8 |
| Combined[a] | 30.3 | 30.9 | 29.4 | 30.2 | 31.1 | 30.6 | 29.2 | 30.5 |

[a] The differences among groups are not significant ($F = 1.59$, $df = 7$ and 152).

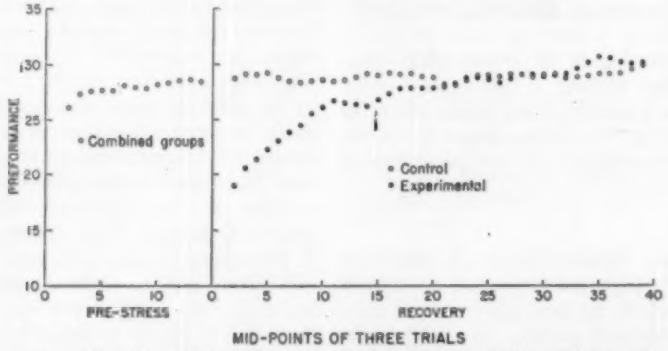


FIG. 2. Curves representing three-point moving averages of performance during prestress and recovery trials.

TABLE 3
MEAN PERFORMANCE SCORES: PRESTRESS PHASE

| Replication | Groups | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $C_1$ | $E_1$ | $C_2$ | $E_2$ | $C_3$ | $E_3$ | $C_4$ | $E_4$ |
| 1 | 27.7 | 26.8 | 27.8 | 27.4 | 27.4 | 28.2 | 27.4 | 28.0 |
| 2 | 27.6 | 28.3 | 27.8 | 26.9 | 28.6 | 28.0 | 27.1 | 27.8 |
| Combined[a] | 27.6 | 27.4 | 27.8 | 27.2 | 27.9 | 28.1 | 27.3 | 27.9 |

[a] The differences among groups are not significant ($F = 0.64$, $df = 7$ and 152).

*Recovery*

*Goal setting behavior.* At the start of the fourth phase of the study changes in operating conditions were introduced for the experimental groups by allowing the subjects once again to work freely at the experimental task; control subjects still operated under the standard conditions that had prevailed for them throughout the experiment. Figure 3 shows that the levels of aspiration of the control subjects continued at the same steady state level as during the preceding two phases of the study; for the four control groups the steady state had persisted for 65, 75, 85, and 95 trials, respectively.

On the other hand, removal of the stress conditions introduced a rapid change in the stated levels of aspiration for all the experimental groups. The ascending curve in Figure 3 shows the general trends of the goal setting process. The process resulted in recovery to prestress levels of aspiration, levels which had characterized the subjects' behavior under conditions identical with those now existing and which corresponded to those maintained by the control groups throughout the experiment. The sensitivity of the goal setting process to the change in operating conditions is shown by the rapid shift from a three-point moving average of 16.1 at the end of the stress trials to an average of 20.8 at the beginning of the recovery trials.

Figure 3 is a composite curve constructed by combining scores for subjects in all four experimental groups. Although it shows general trends, it does disguise one kind of information that is clearly apparent when curves are plotted for each experimental group individually, i.e., the duration of the recovery process appears to be related to the number of stress trials imposed during the preceding phase of the experiment. This relation can be expressed numerically in several ways, all of which give similar results. The data we report here were obtained by a simple graphic method using curves plotted for each experimental group individually: the number of the trial on the abscissa corresponding to the first point on the recovery curve that equaled the prestress mean level of aspiration for the group was taken as representing the number of trials for recovery. The data are presented in Table 6. The trend toward increasing number of trials to recover with increasing number of stress trials which is apparent in these data was found to be significant when tested by Jonckheere's (1954) method and by analysis of variance: by both methods $p < .01$. A more detailed examination of the trend suggests that it is not rectilinear, but tends to approach an asymptote between 30 and 40 stress trials. This possibility was examined by $t$ tests between all the various pairs of experimental groups, using data for the combined replications. The results of the tests are given in Table 7 and show that, as inspection of the basic data had suggested, all differences in number of trials to recover were significant except the difference between Groups $E_3$ and $E_4$, the groups with 30 and 40 stress trials, respectively.

We wish to call attention to one other feature of the recovery curve for the experimental groups. Examination of the curve in Figure 3 suggests that there was some oscillation and "overshooting" in the recovery process after the prestress levels of aspiration had been attained. Had we had some earlier indication of when in the recovery trials these phenomena might have occurred, we would have extended the recovery phase with the intention of determining if and when the phenomena disappeared and a final steady state reached.

TABLE 4
"NET DECREASES" IN LEVELS OF ASPIRATION OF EXPERIMENTAL GROUPS: STRESS PHASE

| Groups | Number of Stress Trials | Median Score |
|---|---|---|
| $E_1$ | 10 | 7.5 |
| $E_2$ | 20 | 9.2 |
| $E_3$ | 30 | 10.2 |
| $E_4$ | 40 | 11.7 |

TABLE 5
MEAN "NET SHIFTS" IN LEVELS OF ASPIRATION OF EXPERIMENTAL GROUPS DURING SUCCESSIVE BLOCKS OF TEN TRIALS: STRESS PHASE

| Block | Stress Trials | Groups | | | | Overall Mean |
|---|---|---|---|---|---|---|
| | | $E_1$ | $E_2$ | $E_3$ | $E_4$ | |
| 1 | 1–10 | 8.0 | 8.2 | 7.3 | 6.9 | 7.6 |
| 2 | 11–20 | — | 4.8 | 3.2 | 3.2 | 3.7 |
| 3 | 21–30 | — | — | 1.7 | 3.6 | 2.6 |
| 4 | 31–40 | — | — | — | 0.6 | 0.6 |

*Performance.* A general picture of changes in performance scores during the recovery phase of the experiment is presented in Figure 2. It seems clear that levels of performance of the experimental groups, as well as their levels of aspiration, had been affected by the changed conditions which operated for them during the stress trials. On the other hand, the performance of the control groups appear to have continued at the same steady state level as during the prestress phase of the study. Recovery from the effects of the stress trials occurred in all groups, the general nature of the recovery process having features in common with that of goal setting behavior under identical conditions.

Data for the experimental groups on the number of trials for recovery of prestress levels of performance have already been presented in Table 6. When analyzed in the same manner as were the level of aspiration scores, the performance data show a significant trend toward increasing number of trials to recover with increasing number of preceding stress trials. The trend tends to approach an asymptote between 30 and 40 stress trials. Examination of the recovery curve in Figure 2 also suggests that oscillation and "overshooting" occurred in the recovery process after the prestress levels of performance had been attained.

## DISCUSSION

Before summarizing our results and conclusions we wish to discuss briefly three aspects of the present experiment.

### *Stress*

Stress, as defined operationally in the present study, was induced by controlling the performance of our experimental subjects, without their knowledge, so that they failed to achieve their stated levels of aspiration during a predetermined number of successive trials. Basically this places the subject in an insoluble problem situation and exposes him to the "stress" that characterizes such situations (Russell, 1953). Berkeley (1952) has demonstrated that exposure to such stress not only affects behavior but also produces physiological changes associated with a wide variety of other forms of stress. One of these physiological changes involves the pituitary-adrenal cortical system (Selye, 1946); exposure to stress results in increased adrenal cortical activity, the increase being proportional to the degree of stress imposed (Hoagland, 1947; Pincus & Hoagland, 1943). In Berkeley's experiment subjects were required to perform the same task used in the present study, failure being controlled by preventing a subject, without his knowledge, from attaining a "minimum score." Effects of exposure to such stress were studied in terms of differences between rates of excretion of urinary 17-ketosteroids under prestress and stress conditions. The results showed that the extent of the discrepancy between aspiration and performance was directly related to the rate of 17-ketosteroid excretion, a multiple correlation coefficient of .70 expressing the magnitude of this relation. In our present experiment we have studied goal setting behavior under basically the same conditions as Berkeley; his findings aid in relating our results to the general concept of stress as it has been widely applied not only by psychologists, but also by research workers in other disciplines.

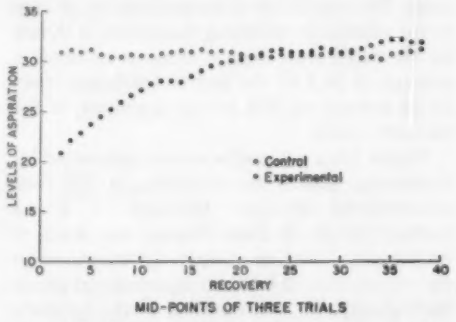


FIG. 3. Curves representing three-point moving averages of levels of aspiration during recovery trials.

TABLE 6
NUMBERS OF TRIALS FOR RECOVERY: LEVELS OF ASPIRATION AND OF PERFORMANCE: COMBINED REPLICATIONS

| Group | Level of Aspiration | Performance |
|---|---|---|
| $E_1$ | 10.1 | 8.4 |
| $E_3$ | 16.7 | 15.2 |
| $E_2$ | 21.4 | 22.0 |
| $E_4$ | 24.9 | 22.2 |

*Lower Limits to Changes in Level of Aspiration under Stress*

During the stress trials the trend in stated levels of aspiration was to decrease consistently as the number of stress trials increased. The decrease was at a decelerating rate, as if the goal setting process were approaching a new steady state level despite continued exposure to stress. Results of other studies (see Lewin et al., 1944) have suggested that there may be absolute limits for any situation, within which the goal setting process may function but beyond which levels of aspiration will not be set. Such limits would presumably be dependent upon other variables affecting the behavior system, e.g., age and sex of the subjects, their socioeconomic status, past success and failure, intelligence, etc. Our experiment was designed to control rather than to study these variables; therefore, we can provide no information on how they may have determined a lower limit for levels of aspiration during the stress trials. It is a reasonable hypothesis that the asymptote which the levels appeared to be approaching may have been set by such a lower limit.

*"Overshooting" and Oscillation*

There is a tendency for feedback systems to respond to change by "overshooting" and oscillation before a stable level is established. Our results give some evidence that these phenomena occurred during the recovery phase after prestress levels of aspiration had once again been reached. Overshooting was the most prominent of the two; it suggests an analogy between the effects of stress on goal setting behavior in our experiment and the "general adaptation syndrome" described by Selye (1946) after measurements of numerous bodily reactions to quite different forms of stress. The analogy is supported by Berkeley's (1952) results which showed that exposure to the same conditions we employed is associated with physiological changes of the kind observed by Selye. Because we had no way of estimating before the experiment began when in the recovery process overshooting and oscillation might occur, we did not extend the recovery phase sufficiently to observe the predicted final establishment of a stable level.

## SUMMARY AND CONCLUSIONS

"Stress" and "goal setting behavior" have been viewed in terms of a model that has been frequently applied to self-regulating physical, biological, and behavioral systems. The model permits certain predictions about the characteristics of goal setting behavior under constant operating conditions and under conditions of stress induced by failure to achieve established goals. The predictions have been tested experimentally using the familiar level of aspiration technique in which feedback, knowledge of results, from the output, performance of a task, affects the behavior system, i.e., establishes or alters levels of aspiration or goals, which in turn influence the future response of the system. Our findings and conclusions may be summarized as follows:

1. The goal setting process very quickly led to the establishment of stated goal levels that remained extremely stable when the conditions under which subjects were operating did not change.

2. The goal setting process was very sensitive to changes in operating conditions. Decreases in stated levels of aspiration were clearly apparent in the very first trials after the introduction of stress; increases in levels were equally apparent immediately following reintroduction of prestress conditions.

TABLE 7
$t$ VALUES[a] OF DIFFERENCES IN RECOVERY TRIALS BETWEEN THE VARIOUS EXPERIMENTAL GROUPS: COMBINED REPLICATIONS

| Group | $E_1$ | $E_3$ | $E_2$ | $E_4$ |
|---|---|---|---|---|
| $E_1$ | | | | |
| $E_3$ | 2.861 | | | |
| $E_2$ | 4.367 | 1.831 | | |
| $E_4$ | 3.892 | 2.099 | 0.353 | |

[a] All are significant except the difference between $E_2$ and $E_4$ ($df = 38$).

3. Under stress conditions there was a significant trend for stated levels of aspiration to decrease consistently as the number of stress trials increased. The decrease occurred, however, at a decelerating rate, as if the goal setting process were approaching a new steady state level despite continued exposure to stress. Results of other studies suggest that this may be dependent upon the existence of a lower limit—determined by other variables known to affect the behavior system, such as age and sex of the subjects, their socioeconomic status, past success and failure, intelligence—beyond which levels of aspiration will not be set.

4. Reintroduction of prestress conditions was associated with an immediate increase in stated levels of aspiration, which continued at a decelerating rate with increasing numbers of recovery trials until a level was reached that was equal to the prestress level and corresponded to the steady state level maintained by control subjects throughout all phases of the experiment. The results picture the goal setting process as following a similar course under both stress and recovery conditions, while affecting level of aspiration in opposite directions.

5. Increasing numbers of stress trials were associated with increasing numbers of trials to recover prestress levels. Under the conditions of the experiment this trend approached an asymptote between 30 and 40 stress trials.

6. Feedback systems tend to respond to change by "overshooting" and oscillation before a new stable level is established. There was evidence that these phenomena occurred during the recovery phase of the present experiment after prestress levels of aspiration had once again been reached. Overshooting was the most prominent of the two.

7. Although our principal interest was in goal setting behavior, we perforce observed the effects of our experimental treatments on our subjects' levels of performance. We were not able to obtain meaningful measures of performance during the stress phase, performance scores being determined artificially by the experimenter. Analysis of the data we did obtain established that levels of performance showed the same general features as levels of aspiration under the same experimental treatments.

We believe that our results provide new information about the process of goal setting, which is frequently involved in normal human reactions to stress. Goal setting has been viewed in terms of a model which we have found useful for integrating results from previous research and for generating predictions, some of which we have put to test in the present experiment.

## REFERENCES


BERKELEY, A. W. Level of aspiration in relation to adrenal cortical activity and the concept of stress. *J. comp. physiol. Psychol.*, 1952, **45**, 443–449.

DAVIS, R. C. The domain of homeostasis. *Psychol. Rev.*, 1958, **65**, 8–13.

FRANK, J. D. Individual differences in certain aspects of the level of aspiration. *Amer. J. Psychol.*, 1935, **47**, 119–126. (a)

FRANK, J. D. Some psychological determinants of the level of aspiration. *Amer. J. Psychol.*, 1935, **47**, 285–293. (b)

HOAGLAND, H. The human adrenal cortex in relation to stressful activities. *J. Aviat. Med.*, 1947, **18**, 450–464.

JONCKHEERE, A. R. A distribution-free *k* test against ordered alternatives. *Biometrika*, 1954, **41**, 133–145.

LEWIN, K., DEMBO, T., FESTINGER, L., & SEARS, P. S. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald, 1944.

MOSIER, C. A., & HALL, J. R. The social grading of occupations. In D. V. Glass (Ed.), *Social mobility in Britain.* London: Routledge, 1954.

PINCUS, G., & HOAGLAND, H. Steroid excretion and the stress of flying. *J. aviat. Med.*, 1943, **14**, 173–193.

ROTTER, J. B. Level of aspiration as a method of studying personality: I. A critical review of methodology. *Psychol. Rev.*, 1942, **49**, 463–474. (a)

ROTTER, J. B. Level of aspiration as a method of studying personality: II. Development and evaluation of a controlled method. *J. exp. Psychol.*, 1942, **31**, 410–42?. (b)

ROTTER, J. B. *Social learning and clinical psychology.* New York: Prentice-Hall, 1954.

RUSSELL, R. W. The comparative study of "conflict" and "experimental neurosis." *Brit. J. Psychol.*, 1950, **41**, 95–108.

RUSSELL, R. W. Behavior under stress. *Int. J. Psycho-Anal.*, 1953, **34**, 1–12.

SELYE, H. The general adaptation syndrome and the diseases of adaptation. *J. clin. Endocrin. Metab.*, 1946, **6**, 117–230.


(Received August 17, 1959)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 389-395

# RIGIDITY IN NEUROSIS AND GENERAL PERSONALITY FUNCTIONING[1]

LAWRENCE A. PERVIN
*Harvard University*

THE phenomenon of rigidity has been of interest to psychologists for a number of years. While a vast amount of research on rigidity has been done, our understanding of the manifold aspects of the construct remains at a minimum. One of the difficulties has been the inconsistency of results obtained by various investigators and by any single investigator. A recent review of the literature by Chown (1959) clearly points up this problem.

As regards the generality of rigidity, the question usually asked is: "If a subject is high on rigidity on one task, will he also be high on another?" The Einstellung Water-Jug Problem is frequently used as a comparative measure in such experiments. Cowen, Wiener, and Hess (1953) designed an alphabet task and Wesley (1953) devised a form to measure manifest rigidity. Both found positive correlations with rigidity on the Water-Jug Problem. While Applezweig (1954) concluded that there is no general factor of rigidity, Eriksen and Eisenstein (1953) and Schaie (1955) (using different tasks) found some correlations. Because of the lack of comparability of results, and the contradictory findings reported, the generality of rigidity remains an unresolved problem. The contradictory results suggest that rigidity may not be a general personality characteristic, but may be limited to particular areas of personality functioning. While it is possible that rigid people are rigid in all aspects of their functioning, it is equally possible that their rigidity is specifically related to certain areas of personality functioning. Such a view was in part responsible for the inclusion of various subtests within measures of intelligence. Thus, while this possibility has been noted elsewhere in the past, the tendency has been to look for broad characteristics with few attempts to look into the more complex aspects of the phenomenon. One part of the present study attempts to examine rigidity in different areas of personality functioning.

Another major point of concern in past research on rigidity has been its relationship to personality variables. In abnormal psychology, rigidity is often used to describe behavioral aspects of patients. Mandl (1954), for example, found paranoid schizophrenics to be more rigid than normals. Fey (1951) found normals to be significantly superior to schizophrenics on the Wisconsin Card Sorting Task. Pullen and Stagner (1953) found that those schizophrenics rated as improved after EST showed a significant decrease in rigidity when compared to those who did not improve.

Of the many studies on rigidity done between the years 1948 and 1958, only three seemed to be directly related to neurosis. Using different tasks, investigators reported contradictory findings. In a qualitative analysis of her data, Angyal (1948) found that rigidity on a perceptual task corresponded to certain clinically neurotic manifestations. Fisher (1950) found neurotics to be significantly less rigid than normals on only a few of the tests given, but most of the results were in the predicted direction. Fisher concluded that no single measure on any task was able to measure general personality trends or predict efficiently. Phillipson (1955) found normals to be significantly more rigid than neurotics.

That so few studies should have been done on neurotic rigidity is surprising in view of the great emphasis that has been put on this phenomenon in the functioning of the neurotic. In referring to the neurotic, Fenichel (1945) speaks of the "repetition of action or attitudes once proved useful that occur whenever similar

[1] This investigation was carried out during the tenure of a Predoctoral Fellowship (MF-9173) from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service. The author would like to express his gratitude to Herbert C. Kelman of Harvard University and Harold Goodglass of the Boston Veterans Administration Hospital. Their guidance was of invaluable assistance throughout all stages of this study. I would also like to thank Frederick Mosteller of Harvard University for his assistance in the statistical analysis.

conditions again prevail or generally, repetition in response to repeating stimuli" (p. 342). White (1956), in referring to early neurotic phenomena, talks of the

> stereotyped, rigid repetition of certain acts in play. They display the earnest, rigid intensity which characterizes neurotic phenomena. The underlying pathology is always a kind of behavior that displays automatic repetition irrespective of the situation, the utility or consequences of the act (p. 237).

From such reports and clinical observations, one may conclude that rigidity is a central phenomenon in the functioning of the neurotic. Therefore, it seems worthwhile to attempt to understand what neurotic rigidity is, and how and why it operates.

The present study examines rigidity in various areas of behavior in neurotics and normals. As far as was possible, an attempt was made to use tasks reported in past research and thereby to provide some basis for comparability of results.

## METHOD

*Subjects.* The subjects were 15 normal, hospitalized patients and 15 neurotic, hospitalized patients taken from the Boston Veterans Administration Hospital.[1] The normals were patients on the general surgical and orthopedic wards. The neurotic population consisted of five obsessive-compulsives, five depressives, and five anxiety reactions. These were based on psychiatric diagnosis. The mean age of the neurotics was 34.1 years, with a range from 23 to 48 years. The mean age of the normals was 33.7 years, with a range from 24 to 45 years. The Vocabulary test of the WAIS was given as a measure of intelligence. The mean IQ of the neurotics was 103.3, with a range from 84 to 129. The mean IQ of the normals was 107.2, with a range from 94 to 123. These means and ranges were considered to be sufficiently similar to justify regarding the normal and neurotic groups as comparable on these two variables.

*Material and Procedure.* In deciding upon the nature of the tasks to be given, the following definition of rigidity was used: Rigidity is the inability to shift set or response when the conditions of the situation require it, i.e., when the existing set or response is no longer rewarding. In all, five tasks were given to each of the 30 subjects. All tasks were similar in that they measured rigidity as defined above. Each was designed to measure a different area of personality functioning. The tasks and areas measured were:

| AREA | TASK |
|---|---|
| A. Problem Solving | A. Einstellung Water-Jug Problem |
| B. Motor | B. Cross-out Letter Test |
| C. Learning | C. Expectancy-Predictive Card Task |
| D. Perception | D. Tachistoscope Recognition Test |
| E. Concept Formation | E. Wisconsin Card Sorting Task (WCST) |

The WCST and Einstellung tasks have been described elsewhere (Fey, 1951; Luchins, 1942). In the WCST two shifts were required. The first concept was color, the second form, and the third number.

The perceptual task consisted of a control series in which five color words were serially shown in a tachistoscope, decreasing speeds until the series had been correctly recognized on two consecutive trials. The speed at which the subject reached the criterion was called the control series recognition speed. For the experimental series, the exposure times were decreased to the original speed of the control series. Here, for every third color word in the series, a noncolor word was substituted. At any one exposure speed, each noncolor word was once substituted for a color word so that, in all, 15 words were shown (five noncolor words and each of the five color words twice). This was done at each exposure speed until two successive correct recognitions of the entire series. All words were equally frequent in usage according to the Thorndike-Lorge Word Frequency List. The words used were brown-bring, white-where, green-great, black-blood, blue-blow. In the perceptual task, rigidity could express itself in two ways: (*a*) in the increase in number of trials until recognition of the mixed color and noncolor series over the number of trials until recognition of the control series and (*b*) in the number of noncolor words seen as color words after the recognition speed of the control (all color) series had been reached or by the time all words should be correctly recognized. If they were not, the subjects could be said to be operating in terms of a set for color words.

In the Expectancy-Predictive Card Task, the subjects were presented with 120 cards. Each card was either black or red. Half the first 30 cards were black and half red, randomly arranged. In the next 60, the ratio of blacks to reds was 3 to 1, again randomly arranged. The last 30 cards were all red. Subjects were asked to predict the color of each card before turning it over. Three measures of rigidity were used in this task:

1. The percentage called black within Cards 36 through 45 and the percentage called black within Cards 46 through 60 were computed for each subject. The difference between these two percentages was then determined for each subject. The greater the difference, representing an increase in percentage called black, the less the rigidity. Thus, rigidity was manifested to the extent that the subject failed to increase the proportion of cards predicted to be black.

[1] I would like to thank the members of the staff and patients at this hospital for their cooperation.

2. The percentage called black within Cards 76 through 90 and the percentage of cards called black within Cards 91 through 105 (all red) were computed for each subject. The difference between these two was determined for each subject. The greater the difference, representing a decrease in percentage called black, the less the rigidity. Thus, rigidity was manifested to the extent that the subject failed to decrease the proportion of cards predicted to be black.

3. In both of the above measures a greater difference represented less rigidity. An overall index of rigidity for this task was computed by combining the above two differences for each subject.

In the motor test, the subjects were asked to cross out every second capital letter and every third "small" letter. The first two lines were all capitals; thereafter, the series of capitals and lower-case letters were intermixed. An error in this test was either a letter not crossed out that should have been or one crossed out that should not have been. The number of errors was taken as the measure of rigidity.

These five tests were given to each subject in a single experimental session. To control for the effect of each task upon the subsequent tasks, the orders of tasks were counterbalanced.

## RESULTS

The results obtained from the rigidity battery were first evaluated by comparing the scores of the normal group with those of the neurotic group, for each of the individual measures of rigidity. Single-tailed $t$ tests are used throughout to determine group differences. A summary of the results of the major measures of rigidity, to be discussed below, is presented in Table 1.

TABLE 1

SUMMARY OF THE DIFFERENCES IN RIGIDITY BETWEEN NEUROTICS AND NORMALS IN VARIOUS AREAS OF PERSONALITY FUNCTIONING

| Area | Normal Mean | Neurotic Mean | $t$ | $p$ Value[a] |
|---|---|---|---|---|
| Problem Solving | 2.40 | 1.86 | .57 | $<.30$ |
| Motor | 2.73 | 4.80 | 1.75 | $<.05$ |
| Learning | | | | |
| Difference between Cards 31–45 and 46–60 | .10 | .03 | 1.23 | $<.15$ |
| Difference between Cards 76–90 and 96–105 | .37 | .19 | 2.52 | $<.01$ |
| Combined Difference | .48 | .15 | 3.46 | $<.01$ |
| Perception | | | | |
| Trials | 5.31 | 6.46 | 2.72 | $<.01$ |
| Words | 2.42 | 4.27 | 1.55 | $<.10$ |
| Concept Formation | 5.47 | 29.47 | 5.11 | $<.001$ |

[a] Single-tailed test of significance.

*Problem Solving.* The mean number of problems solved by the short method were 1.86 for neurotics and 2.40 for normals. The difference was in the predicted direction, but the $p$ value was less than .30. Eight neurotics and seven normals were unable to solve the problem that necessitated the short-cut method. Neurotics were slower on the criterion problems, but since they were slower on all others as well, time was not a discriminating factor between the two groups.

*Motor.* The mean number of errors were 4.8 for neurotics and 2.7 for normals. The difference in means was found to be significant at the .05 level of confidence. Since time could have influenced accuracy, the times for the first two lines (all capitals) and whole test were recorded. The greater the time spent on the first two lines, the less the proportion of time used for the rest of the test. This would represent less difficulty in the shifts that came after the first two lines and consequently less rigidity. The mean percentage of time for the first two lines were .249 for neurotics and .260 for normals. Thus, the difference in errors was not due to an emphasis on accuracy at the expense of time by the normals. (Normals did better, but not because they took more time.)

*Learning.* Using the first measure of rigidity, the mean difference in the two percentages was found to be .027 for neurotics and .094 for normals. The resulting $p$ value was between .10 and .15.

Using the second measure of rigidity, the mean difference in the two percentages was .192 for neurotics and .367 for normals. The difference was found to be significant at the .01 level of confidence.

Using the combined index, the mean combined difference (for the two above measures for each subject) was .148 for neurotics and .477 for normals. The difference was found to be significant at beyond the .01 level of confidence.

The greater difficulty in shifting their probability bases on the part of the neurotics can be seen from the data given in Table 2.

*Perception.* The mean number of trials until correct recognition of the series of color words were 3.5 for neurotics and 3.6 for normals. Thus, there was apparently no difference in perceptual ability per se. The

TABLE 2
DIFFERENCES IN PERFORMANCE AMONG NEUROTICS AND NORMALS ON THE EXPECTANCY-PREDICTIVE CARD TASK

| Cards | Percentage called black (neurotics) | Difference | Percentage called black (normals) | Difference | $p$ Value[a] |
|---|---|---|---|---|---|
| 31–45 | .462 | .026 | .492 | .095 | >.10 |
| 46–60 | .488 | | .587 | | <.15 |
| 61–75 | .579 | | .592 | | |
| 76–90 | .591 | .192 | .675 | .367 | <.01 |
| 91–105 | .399 | | .308 | | |
| 106–120 | .023 | | .031 | | |

[a] Single-tailed test of significance.

mean number of trials for correct recognition of the mixed series (color and noncolor) were 6.4 for neurotics and 5.2 for normals. This difference was significant at the .01 level of confidence. Thus, the neurotics had more difficulty in correctly recognizing the words of the mixed series than did normals. To what was this due? On the second measure of rigidity in this task, for neurotics, the mean number of noncolor words called color once the control series recognition speed had been reached was 4.2. The corresponding mean for the normals was 2.4. This difference represents a $p$ value between .05 and .10. This difference between the groups cannot be accounted for on the basis of some "guessing tendency" because until the control recognition speed had been reached, normals made more such errors than did neurotics.

*Concept Formation.* On the WCST, the mean number of trials until criterion for recognition of the color concept were 5.53 for neurotics and 5.27 for normals. Thus, at this point there would seem to be little difference in what has been called concept formation ability. The first change required was that from color to form. The mean number of trials until criterion for recognition of the form concept were 9.93 for neurotics and 6.87 for normals. This difference represents a $p$ value between .20 and .25. The second change that was required was that from form to number. The mean number of trials until criterion for recognition of the number concept were 29.47 for neurotics and 5.47 for normals. This difference was significant at the .001 level of confidence. When the nature of the responses is further analyzed, it is seen that the neurotic mean for perseverative responses is 27.2 while the corresponding mean for normals is 5.0.

*Intercorrelations of Tasks.* Following the above analysis the intercorrelations of the various measures of rigidity were studied for both normal and neurotic populations. The Pearson product-moment correlation coefficients are represented in Table 3. As can be seen, the correlation coefficients are generally low. Out of the 20 correlation coefficients only 1 reached significance, about what would be expected by chance. Furthermore, the correlation coefficients are different for the two groups. The correlations of the tests with age and IQ were also computed. These correlation coefficients were generally low and somewhat inconsistent.

In the initial conceptualization of the problem it was suggested that each test had something in common with the other tests, which may be called general rigidity, but also had a rigidity specific to it, depending upon the area being represented within the test. It was therefore decided to try to get an estimate of the rigidity in each test that is common to the rigidity in the other tests. In other words, to find out how much of each test is common to all of the other tests. This analysis was also used in an attempt to explain the inconsistency of the correlations noted in Table 3. If, as in Table 3, each test is correlated with any one test, the uniqueness of the two tests is emphasized. However, if each

TABLE 3
PRODUCT-MOMENT CORRELATION COEFFICIENTS OF TESTS OF RIGIDITY FOR NORMAL AND NEUROTIC POPULATIONS

| | WCST | Cross-out Letter Test | Water-Jug Problem | Word Recognition Test |
|---|---|---|---|---|
| Normals | | | | |
| E-P Card Task | .10 | −.01 | .24 | .14 |
| WCST | | .16 | .59 | .11 |
| Cross-out Letter Test | | | .16 | .29 |
| Water-Jug Problem | | | | .18 |
| Neurotics | | | | |
| E-P Card Task | −.09 | −.14 | −.32 | −.32 |
| WCST | | .44 | .22 | .35 |
| Cross-out Letter Test | | | .33 | −.02 |
| Water-Jug Problem | | | | .24 |

test is correlated with the sum of the other four tests, the rigidity common to the set is emphasized. With the latter measure less emphasis is put upon the particular characteristics of the population involved, i.e., in which areas the subjects are rigid. It was hoped that by doing this the correlation coefficients would remain fairly stable from group to group. As a reasonable estimate of what the tests have in common, each test was correlated with the sum of the other tests by assuming equal weights and using the following formula (for Test 5, for example):

$$r_{y5} \sum_{1}^{4} yi = \frac{r_{15} + r_{25} + r_{35} + r_{45}}{\sqrt{4 + 2(r_{12} + r_{13} + r_{14} + r_{23} + r_{24} + r_{34})}}$$

For each of the tests the rigidity measured may be different, but it is felt that this difference is slight, i.e., there is enough in common to call them by the same term. The reasoning behind this is: Each task is correlated with the sum of the other four. For each of the five tasks the composition of the sum with which each is correlated is slightly different. But, three of the four tasks composing the sum will be the same for any two correlations. To assess what is exactly common to the set of five variables requires a more extensive analysis than is given here.

The correlation coefficients for each test of rigidity with the sum of the other four tests of rigidity are given in Table 4. The correlation coefficients in Table 4 indicate that each test measures, in part, a factor common to the other tests. While the correlation coefficients for the normal and neurotic groups are fairly similar, they are not as similar as would have been expected. An examination of the correlations presented in Table 3 indicates that this

TABLE 4

CORRELATIONS OF EACH TEST OF RIGIDITY WITH THE SUM OF THE OTHER FOUR TESTS OF RIGIDITY FOR NORMAL AND NEUROTIC POPULATIONS

| | Normals | Neurotics |
|---|---|---|
| E-P Card Task | .18 | −.12 |
| Water-Jug Problem | .49 | .22 |
| WCST | .43 | .25 |
| Cross-out Letter Test | .23 | .29 |
| Word Recognition Test | .11 | .11 |

TABLE 5

CORRELATIONS OF EACH TEST OF RIGIDITY, EXCLUDING THE E-P CARD TASK, WITH THE SUM OF THE OTHER THREE TESTS OF RIGIDITY FOR NORMAL AND NEUROTIC POPULATIONS

| | Normals | Neurotics |
|---|---|---|
| Water-Jug Problem | .39 | .33 |
| WCST | .37 | .44 |
| Cross-out Letter Test | .25 | .31 |
| Word Recognition Test | .24 | .23 |

is mainly because the correlations of the Expectancy-Predictive Card Task with the other tasks are different for normals and neurotics. This task correlated negatively with the other tasks for neurotics and predominantly positively with the other tasks for normals. If this task is eliminated and each of the tests is correlated with the sum of the other three, the resulting coefficients are as given in Table 5. As can be seen, the resulting correlation coefficients are similar for the two groups. This would indicate that the four tasks have something in common and there are no large differences in what they mean for normals and neurotics.

## DISCUSSION

The results presented above would appear to be consistent with and support the clinical observations that rigidity is a characteristic of neurosis. It would seem that the nature of the control population would tend to further emphasize the differences. These patients were also ill and hospitalized. Furthermore, the mental health of some of the "normal" patients could be questioned. One of the control patients said that he had been discharged from the army because of bad legs, but had received a Rorschach before leaving. Two of the patients in the control group had ulcer conditions. Also, many of the neurotic patients were discharged shortly after testing. Thus, the sampling problem may have led to a reduction of differences.

While four of the five tests resulted in significant differences, the difference on the number category of the WCST is probably the most dramatic. A number of possible explanations for this result may be given: it may be that the number concept per se is particularly difficult for neurotics; it may be

that because a second change is required, the number concept was particularly difficult for neurotics. While other explanations are possible, those given can be studied by rearranging the order of concepts so that number is either the first or second concept. It may be, however, that rigidity becomes particularly evident in a difficult, challenging, or ambiguous situation. Thus, perhaps it would also be of value to use another difficult concept. This could be done by extending the WCST to include another concept (such as shading) or by requiring a concept such as color and form together (allowing some not to be matched at all).

In view of the clear differences found between the two groups, how might the role of rigidity in neurosis be conceptualized? If it may be granted that humans strive for control of their environment, that control is partly dependent upon knowing what to expect, and that one aspect of new and ambiguous situations is that in them one does not know what to expect, then rigidity in neurosis may be seen as operating as follows: Like all humans, the neurotic strives for control of his environment. However, the neurotic has less control over his environment than does the normal person, and experiences his accordingly. Therefore, situations which call into question his ability to master the environment are a particular problem for the neurotic. New and ambiguous situations create such a problem. While normals are usually able to confront such situations and attempt to develop new patterns of behavior appropriate to them, the neurotic flees from the situation in which his ability to control and master is challenged. He interprets the new in terms of the old and therefore feels that old patterns of behavior are still appropriate and adequate. As in the Expectancy-Predictive Card Task, he has difficulty in changing his expectancies. By expecting the future to be similar to the past, he feels justified and adequate in using old patterns of behavior. Rigidity is useful to the neurotic in that it assists him in coping with new and ambiguous situations which would otherwise challenge his doubtful ability to cope with the environment.

It has been demonstrated that the correlations between different tasks measuring rigidity are generally low and vary from group to group. Rigidity is not a general personality characteristic. Within the framework of the present study, the results indicate that a person may be rigid in one area of personality functioning and not in another. If this is true, then it vastly complicates our attempts to study personality traits and characteristics.

The analysis of the correlations for each task with the sum of the other tasks suggested that the Expectancy-Predictive Card Task measures something different for the two groups. For one group this test could be measuring rigidity in the sense of not attending to new information, while for the other group it could be measuring rigidity in the sense of not changing responses even though new information has been received. At this point, no definitive answer to this problem can be given.

Finally, the results indicate that a neurotic may be more rigid than a normal in one or more areas of personality functioning, but not necessarily in all. This suggests the possibility that the particular area or areas within which the neurotic is rigid is of some significance. The results within the neurotic population suggest that different kinds of neurotics (obsessive-compulsive, anxiety, depressive) are high in rigidity on different tasks. While the number of cases within each group (five) is too small to report confidently on these differences, some trends may be noted. There were indications that the anxiety neurotics were particularly rigid on the motor tasks, while the obsessive-compulsives were comparatively more rigid on the more cognitive tasks. The particular areas in which a group of neurotics are rigid may be peculiarly relevant to the nature of their problem areas or ways of handling their problems. It would appear that further investigations into within-neurotic differences might prove helpful in understanding the disorders.

## SUMMARY

The present investigation reported on rigidity in various areas of behavior in neurotics and normals. Each of the 15 hospitalized normals and 15 hospitalized neurotics took five tests of rigidity. Each test was designed

to test rigidity in one of the five following areas: problem solving, motor, learning, perception, and concept formation. Neurotics were found to be more rigid than normals on each of the five tests. Four of the five tests of rigidity resulted in differences between the two groups which were significant at the .05 level of confidence or better.

The intercorrelations of the tests of rigidity were generally low for both the normal and neurotic populations. It was suggested that individuals may be rigid in one area of personality functioning and not in another.

A partial theory of rigidity in neurosis was given along with some suggestions for further research.

## REFERENCES


ANGYAL, ALICE F. The diagnosis of neurotic traits by means of a new perceptual test. *J. Psychol.*, 1948, **25**, 105–135.

APPLEZWEIG, DEE G. Some determinants of behavioral rigidity. *J. abnorm. soc. Psychol.*, 1954, **48**, 224–228.

CHOWN, SHEILA M. Rigidity—a flexible concept. *Psychol. Bull.*, 1959, **56**, 195–223.

COWEN, E. L., WIENER, M., & HESS, J. Generalization of problem-solving rigidity. *J. consult. Psychol.*, 1953, **17**, 100–103.

ERIKSEN, C. W., & EISENSTEIN, D. Personality rigidity and the Rorschach. *J. Pers.*, 1953, **21**, 386–391.

FENICHEL, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.

FEY, ELIZABETH T. Performance of young schizophrenics and young normals on the WCST. *J. consult. Psychol.*, 1951, **15**, 311–319.

FISHER, S. Patterns of personality rigidity and some of their determinants. *Psychol. Monogr.*, 1950, **64**(1, Whole No. 307).

LUCHINS, A. S. Mechanization in problem solving. *Psychol. Monogr.*, 1942, **54**(6, Whole No. 248).

MANDL, BILLIE SUE T. An investigation of rigidity in paranoid schizophrenics as manifested in a perceptual task. Unpublished doctoral dissertation, Purdue University, 1954.

PHILLIPSON, L. K. Personality ridigity as a distinguishing factor between normal and neurotic behavior. Unpublished doctoral dissertation, Purdue University, 1955.

PULLEN, M. S., & STAGNER, R. Rigidity and shock therapy of psychotics. *J. consult. Psychol.*, 1953, **17**, 79–86.

SCHAIE, K. W. A test of behavioral rigidity. *J. abnorm. soc. Psychol.*, 1955, **51**, 604–610.

WESLEY, ELIZABETH. Perseverative behavior in a concept-formation task as a function of manifest anxiety and rigidity. *J. abnorm. soc. Psychol.*, 1953, **48**, 129–143.

WHITE, R. W. *The abnormal personality.* New York: Ronald, 1956.


(Received August 18, 1959)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 396-401

# TWO CORRELATES OF SYMBOLIC SEXUAL AROUSAL: LEVEL OF ANXIETY AND LIKING FOR ESTHETIC MATERIAL[1]

MICHAEL A. WALLACH
*Massachusetts Institute of Technology*

The present research tests two hypotheses derived from Freudian theory concerning the symbolic expression of sexual arousal. The first holds that symbolic expression of sexual arousal is required to a greater extent in persons whose level of anxiety is high, since such individuals are more likely to inhibit the direct or manifest expression of sexual arousal. This hypothesis follows from Freud's writings on dreams (1953a). Relationships between symbolic and manifest projections of sexual arousal have been studied by Clark and Sensibar (1958), and evidence that symbolic sexual arousal in fantasy production permits the expression of impulses denied an overt outlet has been presented in Wallach and Greenberg (1960).

The second hypothesis to be tested states that degree of liking for esthetic materials correlates positively with the degree to which symbolic expression of sexual arousal occurs in response to the materials. This prediction follows from Freud's (1953b) conception of art as a "sublimation" of sexual energy. Proposals toward the study of motivational variables in esthetics have been made by Wallach (1959), and an experiment in this direction concerning the motivational functions served by aspects of graphic expression has been reported by Wallach and Gahm (1960).

While a relationship between degree of liking and degree of symbolic sexual arousal was not found for jazz music in the Wallach and Greenberg study, it seemed very possible that jazz was an inappropriate medium for testing the hypothesis, because degree of preference for jazz may primarily reflect determinants other than an evaluation of the music as such. Preferences regarding jazz may be importantly influenced by judgments of jazz associated activities and things, such as dancing, drinking, and night clubs. It therefore seemed advisable to obtain preference judgments that might reflect evaluation of the music as such more than was the case with jazz. In the present study, degree of liking for certain "classical" selections—music in the Debussy-Ravel idiom—was used for this purpose.[2]

[1] This investigation was supported by a research grant (M-2269) from the National Institute of Mental Health, United States Public Health Service, conducted under the auspices of the Age Center of New England, Inc. Grateful thanks are due Carol Greenberg, Michael Humphreys, Kermit Roosevelt, and Charles Stewart, for aid in data analysis; Donald S. Dunbar and Thomas F. Pettigrew, for making classes available; and Roger Brown, Nathan Kogan, and Lise Wallach, for critical comments on the manuscript.

## METHOD

### *Experimental Subjects*

The subjects were undergraduate women in introductory psychology classes at Simmons College, 73 of whom participated in the story writing and preference rating sessions. Sixty of these subjects also took part in the session at which anxiety was assessed. Sessions of the experiment were administered during class periods without prior announcement. Subjects identified themselves by name on all procedures.

### *Experimental Sessions*

*Story session.* The first class period was devoted to the presentation of three musical selections, with the request to write a story suggested by the piece of music just heard. The procedure was described as a test of creative imagination, and subjects were asked to let the mood induced by the music determine the story that they wrote. A 4-minute period for story writing was provided after the playing of a selection, the subject being permitted to write notes during the music if desired, but not to start the story itself until the music had ended. A blank sheet was in front of the subject while listening to each selection, and each story was written on a separate page. These stories provided the raw material from which symbolic sexual arousal was assessed. In this and all other sessions, a male experimenter presided, introduced as a professor from another university.

*Preference session.* During a second class period, about a month later, the same subjects were asked to make a judgment of degree of liking for each of the three pieces. After hearing each selection again in the same order used in the story session, they were to

[2] The term "classical" is used here in a nontechnical sense as shorthand for referring to the Debussy-Ravel selections presented, in contrast to the jazz selections of the previous study by Wallach and Greenberg (1960).

circle a number on a seven-point scale labeled "strongly dislike" at 1 and "strongly like" at 7. It was emphasized that there were no right or wrong answers.

*Anxiety assessment session.* Level of anxiety was assessed from the Neuroticism scale of the Maudsley Personality Inventory (Eysenck, 1956; Jensen, 1958). Jensen notes that for American university students, the Neuroticism scale correlates .77 with the Taylor Manifest Anxiety Scale (Taylor, 1953). In addition, social extraversion-introversion was assessed from the Extraversion scale of the same instrument, which has a correlation of .81 (Jensen, 1958) with the Social Extraversion scale of the Minnesota TSE Inventory (Evans & McConnell, 1957). The whole untitled questionnaire was filled out during a class session occurring about a month after the musical preference ratings.

### *Content Analysis of the Stories for Symbolic and Manifest Sexual Arousal*

The categories used for scoring the stories in terms of symbolic sexual arousal have been described in detail in Wallach and Greenberg (1960). These categories were originated by Clark and Sensibar (1958) and Beardslee and Fogelson (1958). They include five variables, each of which is scored for presence or absence in a story: motion (description of movement by the body as a whole or by any part of the body from the shoulders down); rhythm (description of any repetitive action, sound, or adverb phrase); peak (a rise followed by a decline in the level of tension implied by the action); penetration (imagery concerning forceful pushing through a resistant medium); and symbolic content (presence of one or more items of story content which are listed by Gutheil—1951, Ch. 3—as sexual symbols, in his formalization of Freud's—1953a—theory of dreams). Because of its direct associations with music, imagery concerning dancing was not scored.

The interpretation that these variables concern symbolic sexual arousal rests on several grounds. First, Clark and Sensibar (1958) have shown that male subjects sexually stimulated by pictures of nude women project more "symbolic content" (the fifth category defined above) than nonstimulated males, both groups being tested in a classroom situation. Second, Wallach and Greenberg (1960) found that the frequency of occurrence of "symbolic content" tends to covary with the frequency of occurrence of the remaining four categories, and the same result was obtained in the present study. Since the five variables cluster empirically, there would seem to be support for Freud's view that all five are symbolic of sexual arousal. Third, an independent sample of female judges was asked to rate the sexual arousal values of the musical selections used in the present study. It will be seen that these independent ratings of the selections for sexual arousal value closely match the average symbolic sexual arousal scores from stories written to the various selections; the same kind of correspondence also was obtained in the previous study by Wallach and Greenberg.

Besides being scored for symbolic sexual arousal, the stories also were scored for manifest sexual arousal according to the system used by Clark (1955). Each of the following kinds of manifest sexual imagery was scored for presence or absence in a story: primary (some kind of evidence for sexual intercourse); secondary (reference to secondary sexual activity such as kissing); tertiary (reference to courting, dating, etc.). As previously noted, dancing imagery was not scored, contrary to Clark's practice.

Two independent scorers scored the 219 stories of our 73 subjects. As each story appeared on a separate sheet of paper identified only by a randomly assigned number, the scorers worked "blindly," not knowing which stories were written by the same subject, nor which stories were written to a particular piece of music.

### *Prior Selection of "Classical" Pieces and Independent Judgments of Arousal Value*

An introductory psychology class of 29 undergraduate women from Radcliffe College heard the three musical selections, and were instructed to rate each piece on a seven-point scale, from "least arousing" (1) to "most arousing" (7), after hearing the selection, in terms of the degree to which it aroused them sexually. The purposes of this judgment session were to obtain independent explicit evaluations of the cue value of each piece for sexual arousal, and to help set the order in which the three selections might be presented as stimuli for stories to the experimental subjects. It was desirable to present the selections in an order of increasing sexual arousal value, to avoid the likelihood that story imagery stimulated by a high arousal piece would be transferred to a subsequently presented selection lower in arousal value. The tests reported below indicated that one of the three selections was judged less arousing than the others; this selection was therefore presented first to the experimental sample.[3]

The first selection was a passage from the third movement of Ravel's *String Quartet in F Major* (Piece A); the second was a passage from the first movement of the same work (Piece B); and the third was a section from the second movement of Debussy's *Petite Suite* (Piece C). The sections to be presented were taped from records, and their respective lengths were 180, 173, and 125 seconds.[4]

## RESULTS AND DISCUSSION

### *Preliminary Considerations*

*Independent judgments of sexual arousal values.* The independent judgments of arousal value made by the 29 Radcliffe women yielded the following means (higher numbers signifying higher arousal value): 3.24 for Piece A, 4.07 for B, and 4.10 for C. The judged arousal value of Piece A thus seems considerably lower than that for B and for C. The difference in a

[3] This same order of presentation had been used for the independent judges.

[4] The Ravel recording was Columbia ML 5245; and the Debussy, Epic LC 3165.

TABLE 1
MEAN SYMBOLIC SEXUAL AROUSAL TO "CLASSICAL" MUSIC FOR THE EXPERIMENTAL SAMPLE, BY SELECTION AND SCORING CATEGORY
($N$ = 73)

| Selection | Mean symbolic sexual arousal score | | | Test of mean difference between scores | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | B-A | | C-A | | B-C | |
| | A | B | C | $t$ | $p<$ | $t$ | $p<$ | $t$ | $p<$ |
| Motion | 0.25 | 0.51 | 0.52 | 3.48 | .01 | 4.16 | .01 | 0.00 | *ns* |
| Rhythm | 0.05 | 0.18 | 0.14 | 2.50 | .02 | 1.59 | *ns* | 0.84 | *ns* |
| Peak | 0.05 | 0.19 | 0.08 | 1.87 | .10 | 0.62 | *ns* | 1.69 | .10 |
| Penetration | 0.00 | 0.03 | 0.00 | 0.14 | *ns* | 0.00 | *ns* | 0.55 | *ns* |
| Symbolic content | 0.41 | 0.64 | 0.55 | 5.66 | .01 | 2.09 | .05 | 0.53 | *ns* |
| Total | 0.76 | 1.55 | 1.29 | 5.67 | .01 | 3.70 | .01 | 1.67 | .10 |

judge's evaluation for each pair of pieces was determined, and the mean difference across judges for each pair of pieces was tested for significance of deviation from zero. Comparing A and B, the mean difference yields a $t$ of 2.68 ($p < .02$);[6] comparing A and C, $t$ is 2.13 ($p < .05$); and comparing B and C, $t$ does not approach significance.

*Reliability of the scoring system for sexual arousal.* Reliability between the two independent scorers was determined for each of the five scoring categories for symbolic sexual arousal. Since a scorer made 219 decisions of presence or absence in the case of each scoring variable, the reliability measure for each variable was taken as the number of stories for which both scorers agreed in their decision of presence or absence, divided by 219. Expressed as percentages, the values obtained were: 87% for motion, 89% for rhythm, 92% for peak, 97% for penetration, and 80% for symbolic content. In a similar manner, reliabilities were determined for each of the three scoring categories for manifest sexual arousal. The number of agreements divided by 219 was 100% for primary, 100% for secondary, and 97% for tertiary sexual imagery. The definitions of the scoring categories hence were clear enough to yield high agreement. Those cases in which the two independent scorers disagreed were resolved by the decisions of a third independent scorer.

*Cue value of the selections for sexual arousal.* Table 1 presents the mean level of symbolic sexual arousal to each of the three pieces taken separately. Since the same person wrote stories to all three selections, significance tests comparing any two pieces were based on the difference between a person's symbolic arousal scores to those two pieces. On all symbolic arousal scoring categories combined, and for motion, rhythm, and symbolic content taken separately (with a trend for peak in addition), Pieces B and A differ significantly, Piece B being higher. On all categories combined, and for motion and symbolic content taken separately (with a trend for rhythm in addition), Pieces C and A differ significantly in symbolic arousal, Piece C being higher. No differences reach the .05 level for C vs. B. To determine whether these results might depend on story length, three $t$ tests were computed, each evaluating the mean differences in length (number of words) for stories to two pieces: A vs. B, A vs. C, and B vs. C. In each test, the score was the difference in story length for a subject's stories to the two selections in question. None of the $t$ values approaches significance. Thus, symbolic arousal to different pieces varies while the length of stories written to these different pieces does not appreciably change.

As was also found in the Wallach-Greenberg study using jazz selections, there is thus a high degree of consistency between the comparative sexual arousal values of the "classical" selections as rated explicitly by a group of independent judges, and symbolic sexual arousal levels to these same selections as measured by a content analysis of stories written without

[6] All significance levels cited in this study are for two-tailed tests.

awareness of the purpose. It might further be mentioned that no cumulative series effect seems to have been operating with regard to symbolic sexual arousal, since symbolic arousal for the last piece was lower (though not significantly so) than that to the preceding one.

Turning now to manifest sexual imagery in response to the "classical" pieces, the frequency of occurrence of such imagery was very low. While the average amount of *symbolic* sexual arousal per person for all three pieces and all five categories combined was 3.60, the average amount of *manifest* sexual arousal per person for all three pieces and all three categories combined was 0.70: 0.00 primary, 0.08 secondary, and 0.62 tertiary imagery per person for all three pieces. The amounts of each of the three kinds of manifest imagery to the three musical selections did not differ significantly. In sum, the present experimental situation may not have been suitable for eliciting manifest sexual imagery, perhaps because of the constraint of the classroom and the presence of a male professor.

### *Anxiety Level and Symbolic Sexual Arousal*

With respect to the first hypothesis—that degree of symbolic sexual arousal varies directly with level of anxiety—a 2 × 2 analysis of variance was performed using anxiety level and preference level as treatment variables, and overall symbolic sexual arousal as the dependent variable (see Table 2). The symbolic sexual arousal score for each subject in this and all subsequent analyses consisted of the subject's total score for all five symbolic arousal categories combined, across all three musical pieces combined. Treatment groups were constructed by separating the sample into four cells consisting of high-high, high-low, low-high, and low-low combinations on the anxiety and preference variables. The sample was dichotomized as close to the median as possible for each variable.[6] Since the number of subjects in the four cells varied, Snedecor's method of unweighted means was used (Grosslee, 1959; Snedecor, 1956, pp. 385–388). The four variances were found homogeneous by Bartlett's test (Walker & Lev, 1953, pp. 193–194), and the fixed constants model (McNemar, 1955, pp. 304–306) was used in calculating $F$ ratios.

[6] The splits made were 52%–48% for anxiety, and 60%–40% for preference.

TABLE 2

MEAN SYMBOLIC SEXUAL AROUSAL TO "CLASSICAL" MUSIC AND ANALYSIS OF VARIANCE, FOR GROUPS HIGH AND LOW IN ANXIETY AND IN PREFERENCE FOR THAT MUSIC

($N$ = 60)

| | High anxiety | Low anxiety |
|---|---|---|
| High preference | 4.78 | 4.07 |
| Low preference | 3.95 | 1.86 |

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| Anxiety | 1 | 26.13 | 7.62** |
| Preference | 1 | 30.93 | 9.02** |
| Interaction | 1 | 6.40 | 1.87 |
| Within cells | 56 | 3.43 | |

** $p < .01$.

The analysis of variance indicates that degree of anxiety is significantly related to level of symbolic sexual arousal ($p < .01$). To further explore this relationship, the product-moment correlation was computed between level of anxiety and symbolic sexual arousal in the stories, yielding an $r$ of .30 ($p < .03$, $N = 60$). The corresponding partial correlation, holding story length constant, is .28 ($p < .04$), indicating that the original relationship is not appreciably influenced by story length.

A possible artifact involved in the anxiety questionnaire measure concerns a tendency to agree in one's answers to any sorts of questions (Couch & Keniston, 1960). Two lines of evidence suggest that such a "yea saying" tendency is not a significant factor in the present study. Since all the items in the anxiety scale are keyed so that "yes" responses lead to a high anxiety score, and two-thirds of the items in the social extraversion scale, which also was administered, are similarly keyed with respect to extraversion, an agreeing response set would lead to a significant positive correlation between the two scales. The obtained correlation, however, is not significant and is actually negative. Further, if yea saying were responsible for the obtained relationship between anxiety and symbolic arousal, it also should result in at least a moderate relationship between extraversion and symbolic arousal, since the latter scale resembles the

former in having the majority of its items keyed so that "yes" responses yield a high score. The relationship between extraversion and symbolic arousal, however, did not even approach significance ($r = -.01$).

Not only an anxiety effect but also a significant interaction between anxiety level and social extraversion-introversion was obtained in the previous study of symbolic arousal to jazz: anxious introverts showed the highest symbolic arousal, nonanxious introverts the lowest, and extraverts fell between these extremes. The question arises as to why no comparable interaction was obtained in the present study. Perhaps the answer is that jazz, because of the social activities and things typically associated with it, has greater relevance to social extraversion-introversion than does the more "classical" music used here. Some support for this suggestion comes from the finding of a significant relationship between degree of social extraversion and degree of overall preference for the jazz ($r = .33$, $N = 72$, $p < .01$), while no such relationship was obtained between social extraversion and degree of overall preference for the "classical" selections.

Finally, no relationship was found between anxiety and manifest sexual imagery (nor between preference for the "classical" pieces and manifest sexual imagery). It has already been noted, however, that the overall amount of manifest sexual imagery in the present experimental situation was very low.

In sum, a relationship has been demonstrated between anxiety level and symbolic sexual arousal, which confirms the first hypothesis of the study. The causal sequences here, however, need further clarification. On the one hand, the subjects may be equally aroused sexually, and those who are more anxious may be constrained to express their arousal symbolically rather than directly—the initial Freudian view. On the other hand, it is possible that more anxious subjects resort more to sexual symbolism because their overall sexual sensitivity or arousability is greater, their stronger sexual arousability making them more anxious and hence more likely to express their arousal only at a symbolic level. Subjects low in anxiety, in turn, may lack anxiety precisely because their overall sexual arousability is low.

### *Liking for Esthetic Material and Symbolic Sexual Arousal*

The second hypothesis stated that degree of symbolic sexual arousal in response to the "classical" selections varies directly with overall liking for the selections. The preference measure was obtained by summing a subject's preference ratings for the three "classical" pieces.

The analysis of variance in Table 2 indicates that degree of preference for the "classical" selections covaries with symbolic arousal. The preference effect, like the anxiety effect, is significant beyond the .01 level. In addition, preference level for the "classical" pieces correlates significantly with degree of symbolic sexual arousal to them ($r = .32$, $p < .01$, $N = 73$). The partial correlation between these two variables, holding story length constant, is .27 ($p < .03$), indicating no appreciable influence by story length differences. That this relationship is not due to a response set to be extreme in marking rating scales may be seen from an $r$ of .01 ($N = 67$) between the sum of a person's preference scale ratings for the "classical" pieces and for the jazz pieces.

Whether symbolic sexual imagery in the stories reflects different basal levels of arousability apart from the musical stimulus or reflects the specific music heard was not at issue with respect to the relationship between anxiety level and symbolic sexual arousal. It becomes relevant, however, in connection with testing the present hypothesis of a relationship between preference for the "classical" pieces and symbolic sexual arousal to them. It is possible that the significant correlation found between these two variables might also be obtained between preference for the "classical" pieces and symbolic sexual arousal to different stimuli. In fact, no correlation is found, however, between preference for the "classical" pieces and symbolic sexual arousal to *jazz* pieces ($r = -.04$), suggesting that the obtained relationship between preference and symbolic arousal for the "classical" pieces at least partly reflects this "classical" music as such.

As a further test of a relationship between

preference and symbolic arousal, the mean preferences for the three pieces were inspected. Since mean symbolic arousal was lower for A than for B and C, average preference levels might be expected to follow the same trend. Mean degrees of preference (higher numbers signifying greater preference) are 4.64 for Piece A, 5.10 for B, and 6.33 for C. The difference in a judge's evaluation for each pair of pieces was obtained ($N = 73$ in each case). Comparing A and B, the mean difference has a $t$ of 2.04 ($p < .05$); for A and C, $t$ is 3.75 ($p < .01$); and for B and C, the $t$ is 6.30 ($p < .01$). Piece A thus was liked less than B or C. Degree of arousal to and degree of preference for the various selections hence are related. Although both B and C were liked more than A, however, C also was liked more than B. Thus, factors besides symbolic sexual arousal level undoubtedly influence preference.

The basis of esthetic preference is a central question for the philosophy of esthetics, and philosophers (e.g., Hospers, 1946) have proposed various criteria in trying to understand how people assess esthetic worth. Freud's "sublimation" theory would predict a positive relationship between degree of liking for a work of art and the degree to which the work permits symbolic expression of sexual energy, since sublimated expression is supposed to be the main psychological function of art. While symbolic sexual arousal obviously is not the sole correlate of esthetic preferences, the results of the present study are in keeping with the sublimation hypothesis. We cannot yet, of course, be sure that the causal sequence is, as Freud would suggest, from symbolic sexual arousal to preference.

## SUMMARY

The present study tests two Freudian hypotheses concerning the symbolic expression of sexual arousal: that symbolic sexual arousal occurs to a greater extent in persons of high anxiety level, and that degree of liking for esthetic materials is positively related to the degree of symbolic sexual arousal in response to those materials. Both hypotheses were confirmed in an experiment with 73 female undergraduates involving the content analysis of imaginative stories written in response to music. Possible implications for anxiety and for the psychology of art were suggested.

## REFERENCES

BEARDSLEE, D. C., & FOGELSON, R. Sex differences in sexual imagery aroused by musical stimulation. In J. W. Atkinson (Ed.), *Motives in fantasy, action and society*. Princeton, N. J.: Van Nostrand, 1958. Pp. 132–142.

CLARK, R. A. The effects of sexual motivation on phantasy. In D. C. McClelland (Ed.), *Studies in motivation*. New York: Appleton-Century, 1955. Pp. 44–57.

CLARK, R. A., & SENSIBAR, MINDA R. The relationship between symbolic and manifest projections of sexuality with some incidental correlates. In J. W. Atkinson (Ed.), *Motives in fantasy, action and society*. Princeton, N. J.: Van Nostrand, 1958. Pp. 117–129.

COUCH, A., & KENISTON, K. Yeasayers and naysayers: Agreeing response set as a personality variable. *J. abnorm. soc. Psychol.*, 1960, **60**, 151–174.

EVANS, CATHERINE, & MCCONNELL, T. R. *Minnesota T-S-E inventory*. (Rev. ed.) Princeton: Educational Testing Service, 1957.

EYSENCK, H. J. The questionnaire measurement of neuroticism and extraversion. *Riv. Psicol.*, 1956, **50**, 113–140.

FREUD, S. *The interpretation of dreams*. (Originally published 1900) Vols. 4 & 5. London: Hogarth, 1953. (a)

FREUD, S. *Three essays on sexuality*. (Originally published 1905) Vol. 7. London: Hogarth, 1953. (b)

GOSSLEE, D. G. Level of significance and power of the unweighted means test. Paper read at American Statistical Association, Chicago, 1959.

GUTHEIL, E. A. *The handbook of dream analysis*. New York: Liveright, 1951.

HOSPERS, J. *Meaning and truth in the arts*. Chapel Hill: Univer. North Carolina Press, 1946.

JENSEN, A. R. The Maudsley Personality Inventory. *Acta psychol.*, 1958, **14**, 314–325.

MCNEMAR, Q. *Psychological statistics*. (Rev. ed.) New York: Wiley, 1955.

SNEDECOR, G. W. *Statistical methods*. (5th ed.) Ames, Iowa: Iowa State College Press, 1956.

TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

WALKER, HELEN M., & LEV, J. *Statistical inference*. New York: Holt, 1953.

WALLACH, M. A. Art, science and representation: Toward an experimental psychology of aesthetics. *J. Aesthet. art Crit.*, 1959, **18**, 159–173.

WALLACH, M. A., & GAHM, RUTHELLEN C. Personality functions of graphic constriction and expansiveness. *J. Pers.*, 1960, **28**, 73–88.

WALLACH, M. A., & GREENBERG, CAROL. Personality functions of symbolic sexual arousal to music. *Psychol. Monogr.*, 1960, **74**(7, Whole No. 494).

(Received August 24, 1959)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 402–410

# CHANGES IN INTERPERSONAL PERCEPTION AS A MEANS OF REDUCING COGNITIVE DISSONANCE[1]

KEITH E. DAVIS AND EDWARD E. JONES
*Duke University*

Among the problem areas to which Festinger's (1957) theory of cognitive dissonance may be applied, none has been as intensively explored in recent experiments as the area of attitude change. The paradigm of creating in an experimental group high dissonance that may be reduced by alterations of belief or feeling has been applied with reference to such attitudinal objects as toys (Brehm & Cohen, 1959b), consumer objects (Brehm, 1956), the interest value of a particular experimental task (Brehm & Cohen, 1959a; Festinger & Carlsmith, 1959), nonfavored vegetables (Brehm, 1959), membership in a discussion group (Aronson & Mills, 1959), marriage before 23 (Cohen, Terry, & Jones, 1959), and cheating (Mills, 1958). In spite of the heterogeneity of research settings, almost all of these studies have involved the creation of cognitive dissonance by inducing the subject (*S*) to engage in some behavior running counter to his private belief. It has been theoretically proposed and empirically verified that, above the minimum of pressure required to induce the *S* to perform the discrepant behavior, the greater the force to comply the *less* the resulting dissonance and the smaller the tendency to adjust one's private beliefs to support the act of compliance.

If an individual has no alternative but to behave in a fashion running counter to his beliefs, relatively little dissonance is created—it is as if he sees himself as the passive victim of fate. If he sees the possibility of behaving in a different fashion more consonant with his beliefs, and still performs in a manner at odds with his private feelings, more dissonance is created. Following this logic, there has been much explicit interest in the variable, "degree of choice." For example, Brehm and Cohen (1959b) found that children, asked to choose one of either two or four toys to keep, increased their liking of the chosen toy more when it was one of four (high choice) than when it was one of two (low choice). Festinger and Carlsmith (1959) found that the greater the monetary incentive for *S*s to argue the merits of a series of boring tasks, the less were the *S*s own attitudes toward the tasks likely to become more favorable. The greater apparent choice in the low incentive condition seemed to produce greater dissonance regarding the persuasion attempt, and this dissonance was reduced when the *S*s rated the tasks as more enjoyable than *S*s in the high incentive (low choice) condition.

The present experiment is both an extension of the above considerations into the area of person perception and an investigation of an alternative mode of dissonance reduction which has largely been ignored in experimental predictions from dissonance theory: the possibility of withdrawing or canceling out one's discrepant behavior instead of showing accommodating changes in attitude.

A recent paper by Jones and Thibaut (1958) emphasized the importance of interaction context in constructing any sensitive and comprehensive theory of person perception. Specifically emphasized in this paper was the role played by the perceiver in defining the situation and eliciting information from the stimulus person. In situations where one individual (A) has, for whatever reason, behaved toward another (B) in a manner which is discrepant with his private feelings about B, we might well expect the production of dissonance and a consequent change in A's attitude toward B. At least dissonance theory alerts us to this possibility and suggests the conditions under which such attitude change might occur. If Person A is pressured by strong incentives or imperative role demands to behave in an overly warm or overly hostile fashion toward B, we would expect little change in A's private impression of B. If, however, A is less certain about what is expected and perceives alternative ways of responding to B, the same extreme behavior would likely produce corresponding changes in

[1] The present research was made possible by a grant from the National Science Foundation (NSF-G8857).

attitude toward B through the mechanism of dissonance reduction. Thus, the "degree of choice" variable seems quite applicable to situations of interpersonal action and perception.

There are, however, occasions similar to the interaction depicted above where A is operating under restraints or incentives which are only temporarily effective. Thus the role requirements are not immutable and A can behave in a different fashion at a later time. An example of this possibility is the case of the practical joker who insults and derogates only to assure his victim, later on, that his remarks were all in jest. Or perhaps more aptly mentioned is the man with a gruff and hostile manner on initial contact who later reveals glimmerings of his heart of gold (and intended to all along).

These examples suggest a variable of more general relevance to dissonance theory. Without exception, the research attempts to create dissonance by inducing the *S* to act in a way that is discrepant from his beliefs, have treated the *S*'s behavior as irrevocable. *S* reads the unpleasant information, he agrees to copy the random numbers, he propagandizes the "naive" subject, he eats the unpleasant vegetable, etc. In all these cases the *S* sees little prospect of taking back or in any public way canceling out the discrepant behavior. Whether or not *S* actually goes through on his commitment to behave in a certain way, he does not anticipate any simple way of neutralizing the behavioral act. It seems logical to suggest that dissonance reduction through attitude change is dependent on the absence of such anticipations. Thus we would expect little dissonance in forced compliance situations where the unpleasant, alien, discrepant behavior can be withdrawn or neutralized.

One form which this undoubtedly takes in many real life situations is neutralization through exaggeration, parody, or dramatic misplaying (see Goffman's discussion of "realigning actions," 1959, p. 190). The child who says "thank you" to his host in response to his mother's imperious prompting may neutralize his remarks (and preserve his autonomy) through sarcastic inflection. The Korean POW who said "I am sorry I called Comrade Wong a *no-good-son-of-a-bitch*" is also effectively canceling out his compliance by emphasizing the wrong words (see Schein, 1958, p. 322).

A more direct form of cancellation, however, is the anticipation that there will be some later opportunity to take back the behavior, to say "I didn't mean it," or "My fingers were crossed," or "I was playing a joke." It is the opportunity for this latter form of neutralization which was varied in the present experiment.

To give a synopsis of the present experiment, *S*s were cajoled (Choice) or assigned (No Choice) to read aloud a negative evaluation of a stimulus person (SP) who could hear but not see the *S*. Half of the *S*s expected to meet the SP later with the experimenter (*E*), at which time *S*'s role and the deceptions involved would be explained to SP (Anticipation condition). The remaining *S*s were led to believe that such a meeting was impossible and that the SP would not be disabused (Nonanticipation condition). In line with the reasoning presented above, the main experimental prediction was: Since dissonance is greatest in that condition where *S* has the illusion of choice and is not given the opportunity to neutralize his behavior, *S* is more negative in his ratings of SP in the Choice-Nonanticipation condition than in any of the remaining three. It was difficult to make specific predictions for the other conditions except for the expectation that the least dissonance and therefore the more positive ratings would occur in the No Choice-Anticipation condition.

## METHOD

### *Subjects*

Exclusive of *S*s used in pretesting, 52 male introductory psychology student volunteers participated in the experiment.[2] The data of 12 *S*s were excluded from the analyses reported below—6 of these were suspicious that the SP was not as presented, and 6 either balked at the task presented them, or later indicated that they were subjectively in a different condition from that implied by the induction. Suspicion and subjective misclassification were evenly distributed across the four

[2] Forty-five of these *S*s were summer school students at the University of North Carolina; the remainder were Duke summer students. We are greatly indebted to the staff of the Organization Research Group of the University of North Carolina for making their research facilities available, and particularly to Joanne Landau who was of invaluable assistance in securing *S*s.

treatments. Only in the Choice-Nonanticipation condition, however, did any *S* balk at reading the evaluation—3 out of 14 *S*s run in this condition refused the induction attempt.[3] The total number of *S*s providing analyzable data was thus 40, 10 appearing in each of the four experiment conditions.

### *Procedure*

The procedure was administered to *S*s one at a time but *S*s were instructed that they were taking part in an experiment in first impression formation and that their task was to come to a first impression judgment of "another subject in the next room." *S*s were led to believe that *E* had taken elaborate precautions to see to it that a "genuine first impression situation would be created," and that these precautions included the pairing up of *S*s who were not taking the same courses, who were not living in the same dormitories, and who had been instructed to arrive at the experimental laboratory at slightly different times to avoid overlap in the halls, etc. Upon arrival of the *S*, *E* conducted him to an observation room behind a one-way mirror which, however, was closed off by a curtain so that the *S* could not see into the adjacent room. *E* explained that he was interested in studying only the impression created by "what a person says and how he says it, and not in the contribution of personal appearance to the impression created by a person." Then *E* asked each *S* to fill out a self-rating scale composed of 10 pairs of conventional antonyms, including generous-stingy, honest-deceitful, and timid-bold. The *S* indicated his self-rating by placing a checkmark in one of five blocks arrayed from one member of the pair to its opposite.

*E* then left the naive *S* alone in one room on the pretext of going to explain the experiment to the SP in an adjacent room. As soon as he reached the second room, *E* turned on a tape recording which broadcast back to *S*'s room a discussion between *E* and the SP. On the tape, *E* explained the nature of the experiment in the same terms he had used with *S*. *E* asked the SP to fill out a self-rating scale and then requested him to respond orally to a series of standard questions about his background, his family, his current interests, etc. and to three moral dilemmas adapted from the Universalism-Particularism Scale (Stouffer, 1949). SP was instructed that his answers to these questions would serve as information for *S* to use in coming to his first impression judgment of the SP, and that after SP had finished responding to these questions *S* would write down a brief personal evaluation of the SP and read this to him over the loudspeaker. As a rationale for this procedure, the SP was told that this would make the experiment "more interesting for you and allow you to hear an honest evaluation of yourself by someone who doesn't know you."

After these remarks, *E* turned off the tape, supposedly leaving the SP to organize his thoughts about the questions which he was to answer and returned to the naive *S*. While waiting for SP to begin, *E* gave *S* a carbon copy of the questions that SP was to answer and made a few informal remarks designed to clarify the procedure. After a short pause *E* returned to the SP and played the second portion of taped material consisting of SP's answers to the standard set of questions and his resolution of the representative moral dilemmas. Because the stimulus information was identical for all *S*s, and because the focus of this study was on the change scores rather than the *S*s' initial impression, no detailed description of the stimulus materials will be presented. The aim was to convey little about the SP except that he was an average, moderately likeable undergraduate. In his appearances on tape, SP seemed slightly nervous; he halted occasionally, groped for words, and seemed quite involved in creating a favorable yet honest impression of himself.

At the close of the SP's talk, *E* turned off the tape and asked the *S* to fill out a "first impression rating scale" (to be described below). Then followed the rationale for the derogation and the Choice and Anticipation manipulations:

> As you will remember, I said at the beginning of the experiment that we are also interested in how people respond to evaluations of themselves. Particularly we are interested in studying the effect on a person of extreme evaluations of himself—evaluations that are either very flattering and complimentary or negative in content.... What I have done is to prepare in advance two quite general evaluations, one of which is heard by every subject who participates in the other half of this experiment. One of these is quite complimentary and contains statements such as, "You sound like one of the most interesting persons that I have met since I came to Chapel Hill (Duke)," and "I would really like to get to know you much better." The other is negative and contains statements such as, "My overall impression was not too favorable," and "I wouldn't go out of my way to get to know you."

The *S*s were then randomly assigned to one of four treatments involving variations in choice and anticipation of interaction. The various sets of instructions were as follows:

> *Choice.* As far as I am concerned, you may read either of these evaluations; but I would like to ask a favor of you. Surprisingly enough, most of my recent subjects have been choosing to read the flattering evaluation of their person, so that I have not been getting much information about the effect of a negative evaluation on people. If you feel able, I'd like to ask you to read the negative evaluation of your partner. Do you think that you'll be able to do this for me?
>
> *No Choice.* Because we are requiring all of our

[3] One *S* in the No Choice-Anticipation or low dissonance condition was, in a sense, self-selected. Originally scheduled for the Choice-Nonanticipation condition, he refused to read the negative evaluation until a meeting was guaranteed (Anticipation) and before being subjected to considerable further pressure (No Choice). Thus in terms of the *final* induction and his subjective placement on the post-experimental questionnaire, this *S* fell in the No Choice-Anticipation condition. The results are quite unchanged by the inclusion or exclusion of this fortieth *S*. For ease of computation and presentation, this *S* was retained.

*Ss* to read one or the other of these evaluations, we have decided that the only fair way to assign one of these evaluations to you is to make it simply a matter of chance. Thus we have decided in advance that all even numbered subjects would be required to read the complimentary evaluation of their partner and that all odd numbered subjects would be required to read the negative evaluation of their partner. Since you are the seventh subject who has taken part in this experiment, here is the evaluation that we want you to read.

*E* then handed the mimeographed copy of the negative evaluation to all *Ss* and began to talk about it only to interrupt himself in one of two ways:

*Anticipation of interaction.* Of course, you realize that as soon as we have had time to study the effect of your reading of this negative evaluation on him, I shall explain to him that this was not your true opinion of him, and you will have an opportunity to meet him and to say anything that you like about your true impression of him. Of course, you don't have to say anything if you don't want to, but you will meet him as soon as we have gathered our data on his reaction.

*Nonanticipation of interaction.* You realize, of course, that we would like to be able to tell this person that this does not represent your true opinion and that it was all part of our experimental setup, but we have found that we cannot do that and continue to do research on this topic. Therefore, we have made this kind of arrangement to protect your identity. As soon as you finish reading this evaluation to him over the microphone, I shall go into the next room and turn him over to an assistant of mine who will take him to a different part of the building where he will, by the use of interviewing and various psychological tests, get at the effect of this evaluation on him. This session will take at least thirty minutes which will give us plenty of time to finish up this part of the experiment and to see to it that you get out of the building without having to meet him and answer any embarrassing questions. Of course, he will never know your name just as you will never know his.

After *S* had had time to read the negative evaluation twice, *E* connected a microphone to an amplifier and appeared to talk briefly with the SP before turning the microphone over to the *S* for him to read it aloud to the SP. The complete negative evaluation follows below (the italicized portion was always dubbed in in pencil to make *S* feel that the evaluation varied to some extent from SP to SP):

As I understand it, my job is to tell you in all honesty what my first impression of you is. So here goes: I hope that what I say won't cause any hard feelings but I'll have to say right away that my overall impression was not too favorable. To put it simply, I wouldn't go out of my way to get to know you. Maybe I'd change my mind if we could talk together in a more natural surrounding, but from the way you spoke—not so much what you said but how you said it—I'd guess that you have some personal problems that would make it hard for us to get along very well. Your general interests and so on just strike me as those of a pretty shallow person.

To be more specific: *Frankly, I just wouldn't know how much I could trust you as a friend after hearing your answers to those moral questions. You took the easy way out every time.* I guess that I should point out some of the things that you said that made a good impression on me, but that would be a kind of a waste of time since the general impression that I have is not too good. That's all that I have to say.

Before leaving the room to take charge of the SP, *E* handed *S* a completed self-rating form which presumably was the one which had been filled out by the SP. *S* was asked to rerate the SP in an attempt to integrate this new information with the information already provided in the "interview." Thus, the second rating was deliberately divorced from the reading of the derogation and given a plausible justification since psychologists "often must combine information from different sources to formulate their conceptions of personality." The answers indicated on SP's self-rating scales were designed to make him appear as typical as possible. In the first place, the items themselves were chosen because it was felt that everyone would respond to them in more or less the same way. In the second place, SP's self-rating indicated that he had a positive self-picture without, however, being especially arrogant or conceited. *Ss* were specifically instructed that the amount of information that they could gain from such a self-rating was limited by the willingness of the other person to reveal things about himself and that they should not feel constrained to agree with the SP about himself.

*E* then left the room for approximately 4 minutes after which time he returned to administer a written questionnaire designed to check on the effectiveness of the experimental manipulations and an oral questionnaire designed to get at their reaction to the SP's self-rating, to recheck the written questions about experimental manipulations, and to ascertain whether or not any *Ss* were suspicious of the procedure. This oral questioning period was followed immediately by an explanation of the deception and purposes of the experiment. Most *Ss* were quite relieved to learn that no one had heard the negative evaluation and left the experimental room with no apparent signs of tension.

### *Rating Scale*

The main dependent variable was the amount of change in each *S*'s evaluation of the SP from before the derogation to after it. The rating scale was constructed specifically to reflect the *Ss*' impression of the SP, and the *before* and *after* forms of the scale were identical, being composed of five trait clusters—likability, warmth, conceit, intelligence, and adjustment—each of which consisted of four simple declarative sentences, two of which were worded positively and two negatively. For example, the conceit cluster was composed of these sentences: "He strikes me as somewhat arrogant and conceited" (negative), "I think that he is basically a modest, unpretentious person" (positive), "This person seems somewhat distant and

aloof" (negative), and "He seems very humble and self-effacing to me" (positive). The *S*s' ratings were made on a six-point scale reflecting the degree of agreement with each of the 20 sentences. For purposes of the present study, the ratings were algebraically summed to provide an overall measure of favorability—that is, all positive items were summed and this total was added to the inversely scored sum of scores for all negative items.

Evidence indicates that both the scale itself and scale change scores are sufficiently reliable. A split-half correlation on the before scores in which one half was composed of the first positive and the first negative item in each trait cluster yielded a .91 value, corrected by the Spearman-Brown formula. A Kuder-Richardson reliability value of .90 was also obtained on the before scores. For the change scores the reliability was somewhat lower but still quite acceptable: .60 when measured by the split-half technique and corrected by the Spearman-Brown formula, and .71 when measured by the Kuder-Richardson technique.

## RESULTS

### *Validation of Experimental Inductions*

As indicated above, only those *S*s who saw themselves in the condition assigned to them (as reflected in their responses to the post-experimental questionnaire) were retained for analysis. Without exception, all *S*s in the Anticipation conditions indicated a definite expectation of meeting the SP immediately after the experiment. None of the *S*s in the Nonanticipation condition did. Nineteen of the 20 *S*s in the No Choice condition felt that "I had no alternative but to give the evaluation I gave." The remaining *S* in No Choice and three *S*s in the Choice condition felt that they had some freedom but that there was "a great deal of pressure on me to choose the evaluation I gave." The remaining Choice *S*s scattered their responses over three alternatives reflecting greater degrees of subjective choice or freedom to choose. Since there was little or no overlap between the questionnaire responses of the retained *S*s in different conditions on questions relevant to condition differences, the experimental inductions may be considered valid almost by definition. Since only six *S*s excluded themselves by their questionnaire responses, indicating that subjectively they were not in the induced condition, the experimental inductions may be considered basically valid even without regard to *S* selection. It should also be noted, again with reference to the post-experimental questionnaire, that there was no interaction between subjective choice and anticipation. For example, the Choice *S*s in the Anticipation condition felt just as much choice as the Choice *S*s in the Nonanticipation condition.

A final matter of importance concerns the extent to which the evaluative reading was actually seen to be discrepant—in a negative direction—from each *S*'s initial attitude. Except for two *S*s in the high dissonance condition, all *S*s saw the evaluation as less favorable in tone than their own attitude. Most *S*s thought the discrepancy was considerable.

### *Effects of Choice and Anticipation on Rating Changes*

The dependent variable of critical interest is the summary rating change score for each *S*. The means and standard deviations of these change scores are presented for each experimental condition in Table 1. A simple analysis of variance indicates that the difference between all four means is significant ($p < .05$). The variances are quite homogeneous, though there is some tendency for the *S*s in the Anticipation conditions to show greater variability than those in the Nonanticipation conditions.

As always, interpretation of significant differences in change scores depends on the relationship between change scores and before scores. In the present case, the differences between the four before-score means did not approach significance. Also, none of the four

TABLE 1

MEAN BEFORE AND CHANGE SCORES AND STANDARD DEVIATIONS OF SUMMARY RATING EVALUATIONS

| | Condition | | | |
|---|---|---|---|---|
| | Choice | | No Choice | |
| | Antic | Nonantic | Antic | Nonantic |
| Before[a] | | | | |
| *M* | 51.0 | 55.5 | 48.0 | 54.7 |
| *SD* | 8.92 | 17.31 | 11.89 | 15.25 |
| Change[b] | | | | |
| *M* | −1.8 | −7.7[c] | −2.2 | −1.7 |
| *SD* | 6.63 | 4.79 | 5.39 | 3.43 |

Note.—The lower the before score the more favorable the rating. A minus change score indicates change toward a less favorable impression. $N = 10$ for each cell; total $N = 40$.

[a] $F$ betw/with = .641.

[b] $F$ betw/with = 3.41, 3 and 36 $df$, $p < .05$.

[c] Individual mean comparisons with C-Na (Choice Nonanticipation)—NC-Na: $t = 3.222$, 19 $df$, $p < .01$; C-A: $t = 2.278$, 19 $df$, $p < .05$; NC-A: $t = 2.413$, 19 $df$, $p < .05$.

correlations between before scores and change scores differed significantly from zero; two were in the positive, two were in the negative direction. As a final check on the possibility that the change scores were influenced by the before scores in some systematic way differing across treatments, an analysis of covariance was done which indicated that the difference between means was still significant after partialing out any effects of before score.

*The role of choice.* The main hypothesis of the experiment was that the illusion of choosing to read the negative evaluation plus the awareness that there would be no disabusing interaction with the SP would combine to produce the greatest amount of dissonance. The most direct consequence of this heightened dissonance would be a high negative change on the summary evaluation score, relative to the other three experimental groups. It is clear by inspection of Table 1 that this main hypothesis is confirmed. The Choice-Nonanticipation group mean is significantly different from each of the other group means at the 5% level or lower. The remaining three groups are obviously very similar in mean and the differences between them do not approach significance.

It is somewhat surprising to note that the No Choice-Anticipation condition, expected to produce the least dissonance, results in as much negative change as do the two moderate dissonance conditions (Choice-Anticipation and No Choice-Nonanticipation). Apparently, whether or not one has any freedom to read the negative evaluation is not related to the production of dissonance as long as *S* feels he can retract his behavior after the experiment. When this avenue of dissonance reduction (or prevention) is closed, however, the present experiment clearly confirms the role of choice or commitment in the production of dissonance.

*The role of anticipation.* In view of the pattern of rating change results, one could make the same sort of remarks about the role of anticipation as about the role of choice: knowledge that there will be no interaction is essential in the production of dissonance only when *S* has the illusion of some choice in reading the evaluation. However, as we might infer by the tendency toward greater variability in the anticipation groups, the anticipation variable is probably more complex than the choice variable—or at least the lines between conceptual status and operational definition are more tenuous with regard to anticipation than choice. Choosing to behave in a certain manner is a rather direct index of committing oneself to the behavior, of "owning up to it," or acknowledging it is at least partially self-caused.

In the present experiment, the anticipation of meeting or not meeting the SP was manipulated to allow half of the *S*s to reduce or prevent dissonance by anticipating the withdrawal of their behavior rather than changing their attitude. But anticipation had other meanings to these *S*s as well. A few *S*s openly expressed anxiety at the prospect of meeting the SP. As one *S* not too lightly said: "I hope he's not a big guy." There were probably others who shared some of these feelings but did not express them openly. Nevertheless, there is evidence from the post-experimental questionnaire that such anxiety reactions were definitely in the minority. When asked, "How appealing is (or would be) the prospect of meeting the other subject to explain your part in this?" 10 (or half) of the Anticipation *S*s checked the alternative, "very much like to meet him and explain." Six more *S*s checked the next alternative, "rather like to meet him and explain." Only three *S*s checked the third or neutral alternative, with one saying "rather not meet him . . . ." In contrast to this, *S*s in the Nonanticipation condition tended to feel much more neutral about meeting the SP: 11 out of 20 checking the neutral alternative. The *S*s in the Anticipation conditions are, in fact, significantly *more* eager to meet the SP than the *S*s in the Nonanticipation conditions ($\chi^2 = 4.401$, 1 *df*, corrected for continuity, $p < .05$). This suggests, perhaps, a "sour grapes" reaction in the nonanticipation *S*s—since they cannot meet the SP anyway, the dissonance produced by this is partly reduced by expressing a lack of interest in such a meeting. In any event, the relative eagerness of the anticipation *S*s to meet the SP indicates that anticipation of interaction plays a greater part in dissonance reduction than in anxiety arousal, and that the latter reaction is atypical.

*Performance Variations and Attributions of Suspicion*

It was impossible to control the facility and sincerity with which the *S*s read the derogatory evaluation, except by strong instructions urging simulated sincerity and the provision of a brief practice reading period. It is conceivable that some *S*s deliberately tried to sound insincere or to give cues to the SP that their performance was other than genuine. This possibility was mentioned in the introduction as an alternative means of discounting behavior which is discrepant from other established cognitions. For this reason, *E* attempted to rate each performer on the sincerity of his performance. While it turned out to be impossible to make reliable ratings of such an intangible quality, *E* was aware of only one *S* who seemed to be deliberately conveying cues by exaggerating his part over the microphone. Probably because of the constraints inherent in the situation, then, attempts at caricature in the reading performance did not serve as a major means of dissonance reduction in the present experiment.

Nevertheless, *S*s did vary in response to one question which asked whether they thought SP was likely to be suspicious of the sincerity or source of the evaluation which was read to them. Fifteen of the *S*s felt that the SP was likely to be suspicious, and these distributed themselves quite evenly across the different conditions. Since the attribution of suspicion is a possible way of reducing dissonance, it is important to note the relationship between such attribution and rating change scores. Here a rather curious and unexpected finding emerges. In the Nonanticipation conditions, there is no relationship between attribution of suspicion and change score. In the Anticipation conditions, those who attribute suspicion to the SP show significantly greater negative rating change than those who do not ($t = 3.41$, $p < .01$). In none of the comparisons does the choice variable show any relationship to the attribution of suspicion.

The difference within the Anticipation conditions seems paradoxical and quite difficult to explain. One would assume that attributing suspicion to SP would be one way of neutralizing the situation and avoiding dissonance; yet those who attribute suspicion are precisely the ones who show the greatest dissonance reduction by changes in rating. There are a number of equally plausible explanations for this finding which are neither refuted nor supported by available data. Any commitment to a particular explanation would therefore be quite arbitrary. In short, the findings with regard to attribution of suspicion remain essentially a mystery.

*The Role of Self-Rating Discrepancy*

In order to support the rationale of the experiment, it was essential to provide some further information about the SP during the interval between the first and second rating of him. An attempt was made to construct a bogus self-rating performance for the SP which was purposely uninformative and yet could serve as a minimal justification for changes in rating. Some information was provided by this SP self-rating, however, and many *S*s claimed that they used this information in re-evaluating him. Many of those who became more negative in their ratings felt that the SP rated himself in a conceited or egocentric way. Since the attribution of conceit would presumably bear some relation to one's own self-ratings on the same scale, the discrepancy between the *S*'s and the bogus SP's ratings was considered as a possibly confounding variable. If, for example, *S*s in the Choice-Nonanticipation condition were for some reason more modest in their self-ratings than *S*s in the other conditions, they would tend to think of the SP as relatively conceited and therefore dislike him. Such a chance occurrence would make it unnecessary to invoke dissonance theory in order to explain the results.

There were variations in average self-rating discrepancy from group to group, and it so happens that the largest discrepancy occurred in the critical Choice-Nonanticipation condition. None of these differences in self-rating discrepancy means departed significantly from chance, but as a final control over this factor covariance analyses were conducted to partial out the effects of self-rating discrepancy in comparing Choice-Nonanticipation with each other condition in turn.

Suffice it to say that there were no essential changes in results or significance levels except for the overall between groups effect, which drops just below the 5% level. Since this overall comparison includes some groups between which no differences were predicted, and the crucial *t* test comparisons (Choice-Nonanticipation versus the other groups) remain essentially unchanged, we may conclude that the present results cannot be explained by chance variation in the self-ratings of *S*s assigned to different conditions.

## DISCUSSION

The results of the present experiment indicate the conditions under which a person, whose behavior toward another departs from his initial attitude toward the other, changes his attitude to make it consonant with his behavior. The person or actor must (*a*) feel that he had some freedom *not* to behave in the discrepant manner, and (*b*) realize that he cannot easily disclaim the behavior in the eyes of the target person. According to the findings presented above, the combination of these two conditions is necessary and sufficient to produce enough dissonance to motivate attitude change. In the absence of either condition, the amount of attitude change is negligible and significantly less than when both are present. We assume in this general statement that it is merely the amount of discrepancy between behavior and initial attitude which is important and not the direction of discrepancy. Since the present experiment only involved behavior which was more negative than the initial attitude, the generality of the present results is contingent on further empirical investigation.

The first of the two necessary conditions (Choice) has received considerable attention in the research of others. The present experiment thus confirms and extends to a different attitudinal object (a person) the findings of Brehm and Cohen (1959a, 1959b) and of Festinger and Carlsmith (1959). The present experiment indicates, however, that the choice to behave in a certain manner must have fairly irrevocable consequences in order for dissonance and consequent attitude change to appear. When *S*s are led to believe that they will soon meet the target of their negative evaluation, and that they may then explain away their behavior, the choice variable has no effect on attitude change. In considering the consequence of "forced compliance," therefore, it is essential to note the available means of neutralizing the behavioral act involved in the expected dissonance. Can the person easily change the meaning of this act or modify its significance? Can he later disown or revoke the behavior? One effect of such possibilities, specifically the latter one, is brought to light by the present results. The commitment implicit in choosing to behave in a manner discrepant from attitude, must be a commitment to a durable, preferably public performance, and not a series of acts which can easily be neutralized or taken back.

The implications of the present results for theories of social interaction and perception are rather straightforward. Certainly some clues are provided for predicting the effects of interpersonal communication on interpersonal attitude. A hostile comment that is clearly demanded by role requirements (No Choice) should lead to less change in the communicator's attitude than a hostile comment that represents a response to more ambiguous circumstances (Choice). From this we might further infer that the well practiced performer, the man who feels every interpersonal act is "spelled out in the manual," and who knows the pattern of behavior required by each role he is forced to play—such a man would not likely be affected by his own utterances toward others. On the other hand, the improviser, or the person who is uncertain of situational and role requirements, is much more apt to be constantly in the throes of dissonance elevation and reduction. In the process, presumably, his impressions of the same other may actually be quite variable in the course of an extended conversation. The potential dilemma of this latter type of person is blunted by the fact that he may succeed in always behaving toward another in a manner consonant with his initial attitude. Given some (preferably ambiguous) pressure to depart from this consonance of communicative behavior, however, there should be clear fluctuations in attitude or personal impression.

SUMMARY

This study was designed to demonstrate that dissonance produced by inconsistencies between one's evaluation of another and one's behavior toward him would be reduced, under certain conditions only, by changing one's evaluation to conform to the behavior. All *S*s read a negative evaluation about a stimulus person over a microphone presumably broadcasting to him. *S*s were instructed to act as though this was their true evaluation of the stimulus person. Half of the *S*s (Choice conditions) were given some opportunity to read a positive evaluation but urged to read the negative one; the remaining *S*s (No Choice conditions) were simply assigned to read the negative evaluation. Half of the *S*s in each of these groups were in turn informed that they would have a chance to meet the stimulus person immediately after the experiment when all would be explained to him (Anticipation conditions); the remaining *S*s were told that no meeting would be possible and that the stimulus person could not be informed of the deception (Nonanticipation conditions).

The major prediction derived from dissonance theory was that only those *S*s in the Choice-Nonanticipation condition would have sufficient dissonance so that a negative change in the impressional evaluation of the stimulus person would occur. This prediction was confirmed, though it was surprising that the No Choice-Anticipation group showed no less attitude change than the "moderate dissonance" conditions: No Choice-Nonanticipation and Choice-Anticipation. Some incidental findings were presented and the results were discussed in terms of dissonance theory and implications for interpersonal perception.

REFERENCES

ARONSON, E., & MILLS, J. The effect of severity of initiation on liking for a group. *J. abnorm. soc. Psychol.*, 1959, **59**, 177–182.

BREHM, J. W. Postdecision changes in the desirability of alternatives. *J. abnorm. soc. Psychol.*, 1956, **52**, 384–389.

BREHM, J. W. Increasing cognitive dissonance by a *fait accompli*. *J. abnorm. soc. Psychol.*, 1959, **58**, 379–382.

BREHM, J. W., & COHEN, A. R. Choice and chance relative deprivation as determinants of cognitive dissonance. *J. abnorm. soc. Psychol.*, 1959, **58**, 383–387. (a)

BREHM, J. W., & COHEN, A. R. Re-evaluation of choice alternatives as a function of their number and qualitative similarity. *J. abnorm. soc. Psychol.*, 1959, **58**, 373–378. (b)

COHEN, A. R., TERRY, H. I., & JONES, C. B. Attitudinal effects of choice in exposure to counterpropaganda. *J. abnorm. soc. Psychol.*, 1959, **58**, 388–391.

FESTINGER, L. *A theory of cognitive dissonance*. Evanston, Ill.: Row, Peterson, 1957.

FESTINGER, L., & CARLSMITH, J. M. Cognitive consequences of forced compliance. *J. abnorm. soc. Psychol.*, 1959, **58**, 203–210.

GOFFMAN, E. *The presentation of self in everyday life*. Garden City, N. Y.: Doubleday Anchor, 1959.

JONES, E. E., & THIBAUT, J. W. Interaction goals as bases of inference in interpersonal perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford: Stanford Univer. Press, 1958. Pp. 151–179.

MILLS, J. Changes in moral attitudes following temptation. *J. Pers.*, 1958, **26**, 517–531.

SCHEIN, E. H. The Chinese indoctrination program for prisoners of war: A study of attempted "brainwashing." In Eleanor Maccoby, T. Newcomb, & E. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1958. Pp. 311–344.

STOUFFER, S. A. An analysis of conflicting social norms. *Amer. sociol. Rev.*, 1949, **14**, 707–717.

(Received August 31, 1959)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 411-418

# SOME PERSONAL AND SITUATIONAL DETERMINANTS OF YIELDING TO INFLUENCE[1]

AMBALAL S. PATEL[2] AND JESSE E. GORDON
*University of Wisconsin* *University of Michigan*

THE purpose of this study was to discover some of the personal and social variables that are related to yielding to influence in a situation in which subjects (*S*s) are given the option of yielding to or ignoring the influence.

In typical experiments in this area, *S*s are asked to make judgments before and after experimentally produced "false" judgments are presented; changes in *S*s judgments in the direction of the artificial ones are taken as evidence for suggestibility or yielding to influence (Moore, 1921; Sherif, 1936). Studies in which judgments provide the dependent measures, such as literary judgments, judgments of line lengths or of extent of movement in the autokinetic effect, yield evidence that many *S*s modify their verbal behavior to conform to group norms or to the prestige suggestions. However, such studies provide little evidence that *S*s do in fact change their attitudes sufficiently for there to be a change in their overt behavior outside of the experimental situation. It is possible that *S*s, usually both curious and uncomfortable in experiments in which they participate, were in the position of searching not for influence but for guidance in responding appropriately. The experimenter produced cues, which are usually labeled as prestige suggestions or social influence, may be perceived by *S*s as guides or cues for structuring the task and thus reducing *S*s' trial and error. Operating on such an assumption, *S*s may give verbal responses of a conforming nature in order to avoid censure, without there being any more generalized effects of the influence on their judgments or response tendencies outside of the situation. When Hamlet leads Polonius to agree that a cloud has in turn the shape of a camel, a weasel, and a whale, we understand that he is extracting a response of expediency and not necessarily producing a change in Polonius' perceptions of clouds.

To an extent this problem is related to the deliberate subtlety with which the influence is introduced in the experiments. The situations are kept vague so that *S*s will not discover the purpose of the experiment and will not know that they are being subjected to influence. This vagueness is made necessary by the experimenter's (*E*'s) fears that should *S* know that he is being subjected to influence, he would resist it. Such experiments do throw considerable light on the extent to which a person's verbal behavior may be controlled by the cues present in a situation, but it cannot be said that *S*s are yielding to influence or accepting suggestions when, as far as *S* is concerned, no suggestions have been offered.

An experiment by Coffin (1941) corrects this situation to some extent, by making the influence explicit. He asked his *S*s to solve mathematical problems, and penciled in both correct and incorrect "hints" for the solutions of the problems. However, even here *S*s were not free to refuse the suggested hints, because Coffin announced to his *S*s that he had given hints in order to make the problems easier to solve. Assuming the rapport that exists between students and a teacher-researcher, it would seem that Coffin's *S*s could better be described as following instructions, both explicit and implicit, when they used the "hints" given them, than as yielding to influence. It is therefore not surprising that those *S*s with little knowledge of mathematics employed more hints: they did not have the knowledge necessary to recognize that the hints were misleading in some cases, and considering their relationship with the teacher-experimenter, assumed that the hints were valid guides to the solutions of the problems.

[1] This paper was based on a thesis which was submitted in partial fulfillment of the requirements for the MS Degree in the Department of Education of the University of Wisconsin.

The authors wish to extend their appreciation to A. H. Edgerton of the Department of Education and D. A. Grant of the Department of Psychology for their assistance in this study, and to the Wisconsin Alumni Research Foundation for its financial aid in support of this research.

[2] Now at the University of Baroda, Baroda, India.

This problem is a part of the relative importance of normative and informational influence. These terms, proposed by Deutsch and Gerard (1955), identify different aspects of suggestions: a suggestion is informational when it transmits some presumable trustworthy data about the to-be-judged stimuli to an *S*; a suggestion is normative when it represents a response that is characteristic of a group. It is obviously quite difficult to separate these aspects from each other. A suggestion that is informational usually comes from a source, thus bringing with it the implication that it carries the authority associated with the prestige of the source person and the group to which he belongs, and is thus normative. In general, the informational value of a suggestion varies with *S*s' state of knowledge; the normative value varies with the relationship between *S* and the source of the suggestion. Deutsch and Gerard (1955) attempted to separate these aspects of suggestion by having some *S*s report their judgments anonymously, after having been subjected to influence, while others made their judgments with varying degrees of public commitment to the influencing group. They found some modification of judgment even in the anonymous group, in which normative influence was presumably not operating, and suggest that this modification is a result of informational influence. However, there was no effort to manipulate the state of *S*'s own information regarding the stimulus; without such manipulation, it is not possible definitely to identify the influence as informational in the anonymous group. It is possible that even though *S* was not publicly committed to his judgment, his desire to be like the "stooges" may have resulted in the modification of his judgments. If this were the case, then the suggestions should be considered normative rather than informational. Samelson (1957) tested and confirmed the hypothesis, derived from this distinction between normative and informational influence, that where *S* was given reason to doubt the accuracy of the source (thus reducing its informational value), conforming behavior would be reduced. These two studies thus provide evidence that both informational and normative suggestions are capable of modifying judgment, although Deutsch and Gerard were not able to manipulate normative suggestions without informational components being included.

The study reported here throws some light on the relationships among informational and normative influences and acceptance of influence in a situation in which the interactions of these influences may be observed. The informational value of influence is manipulated by varying the initial level of information of the *S*s relative to the task; the normative value of the same influence is manipulated by ascribing the influence to high and low prestige sources. Under these conditions, if it is hypothesized that both aspects of influence produce modifications in *S*s' behavior, then high prestige influence directed at the low knowledgeable *S*s should produce high acceptance of the influence, and low prestige influence directed at the high knowledgeable *S*s should produce little acceptance of the influence. *S*s are made explicitly aware that suggestions are present, and that they may choose to accept or reject them. The procedures were designed to eliminate the implication that the suggestions presented were necessarily informational, rather than normative, and that they constitute a valid guide for behavior in the experiment.

The experimental task facing the *S*s in this research was to indicate the correct synonyms for English words, the stimulus words being graded in difficulty. Because the test was administered in high school English classes by the English teachers,[3] and was not introduced as an experiment, it seems likely that *S*s' responses to the test items, whether or not they accepted the suggestions given, represented their honest belief that the words they checked as synonymous were in fact so. It is doubtful that their choices were influenced by the kind of expediency illustrated by Polonius' responses to Hamlet, and which may have been operating in "group pressure" experiments. It is more likely, therefore, that *S*s' acceptance of suggestion or influence produced changes not only in their verbal responses, but also in their cognitive functioning related to the influenced responses.

The independent variables whose relation-

[3] The authors wish to acknowledge the assistance and cooperation of the administrators and teachers of the Portage High School, Portage, Wisconsin.

TABLE 1
ANALYSIS OF VARIANCE OF SUGGESTIBILITY SCORES

| Source | *SS* | *df* | *MS* | *F* |
|---|---|---|---|---|
| Main Effects: | | | | |
| Grade | 13837.10 | 2 | 6918.55 | 17.28*** |
| Prestige | 1928.01 | 1 | 1928.01 | 4.81* |
| Sex | 2821.47 | 1 | 2821.47 | 7.05** |
| Difficulty | 14890.30 | 1 | 14890.30 | 37.18*** |
| Interactions: | | | | |
| Grade × Prestige | 3659.40 | 2 | 1829.70 | 4.57* |
| Grade × Prestige × Sex | 3614.02 | 2 | 1807.01 | 4.51* |
| Grade × Prestige × Difficulty | 4405.39 | 2 | 2202.69 | 5.50** |
| Grade × Prestige × Sex × Difficulty | 4752.92 | 2 | 2376.46 | 5.03** |
| Error: | | | | |
| Within Groups | 48054.69 | 120 | 400.46 | |
| Total: | 98139.78 | 143 | | |

* $p < .05$.
** $p < .01$.
*** $p < .001$.

ships to the acceptance of suggestions were measured were: prestige of the source of influence, ambiguity of the problems for which suggested solutions were provided, sex of *S*, and age-grade of *S*, varied independently of ambiguity of the problems.

## METHOD

*Ss* were 72 high school students, from the tenth, eleventh, and twelfth grades. Both sexes were represented in each grade.

*Ss* were given Lepley's Synonym Vocabulary Test (Lepley, 1955),[4] consisting of eight parts. Each part contains 30 stimulus words and 5 response words from which *S* must pick the correct one for each stimulus word by marking A, B, C, D, or E on an IBM answer sheet. Correctness of response is in terms of similarity in meaning to the stimulus word. The only differences among the eight parts are in the stimulus and response words, and their difficulty as determined from Lepley's standardization groups.

Suggestions were given *Ss* by marking in pencil in the questionnaire booklets an incorrect letter next to 50 of the stimulus words, and a correct letter next to 11 of the stimulus words. The first 4 stimulus words marked were marked correctly, in order to create confidence in the markings by the *Ss*. Thereafter, in the test, the remaining 7 correct marks were randomly distributed among the 50 incorrect ones. The markings were identical for all *Ss*. Only the incorrectly suggested items were used for the analysis of the effects of suggestion.

Use of the marked "hints" was made optional for the *Ss* by telling them that they were to put their answers on the IBM answer sheets and make no marks on the questionnaire booklet. The teacher who administered the tests then indicated that the test had been taken by other classes previously, and she noted that some of the students in the other classes had put marks on the test booklets, but that these marks should be ignored. Thus the suggestions were directly provided to the *Ss*, who were explicitly aware of the suggestions, at the same time as they were permitted to accept or reject the suggestions without being subjected to pressure to accept from the supposed originator of the suggestion.

Prestige of suggestion source was manipulated by giving differing instructions to half of each grade group. In one half of each group, the teacher indicated that the class which had taken the test previously was one year lower in school than the *S* group; in the other half, it was indicated that the class which had taken the test was one year ahead of the *S* group. Thus half of the tenth grade sample was told that the test was taken just previously by ninth grade students (who were, by implication, the ones who made the marks on the test booklets), and half were told that it was taken by the eleventh grade.

Ambiguity of the problem facing the *Ss* was manipulated independently of grade level by having each *S* rate the words whose synonyms had been "suggested" several days after the test was given. Each *S* was given a list of the 61 marked stimulus words and two others as an example, and was asked to place a cross next to those which he considered to be difficult to understand in general, or whose meaning he did not know, and a check next to those whose meaning *S* knew. *S*'s scores for acceptance of influence were computed individually as the proportion of his "easy" and of his "difficult" words on which he took the incorrect suggestions.

Analysis of the data was in a 3 × 2 × 2 × 2 factorial design, consisting of three levels of age-grade status, two sexes, two levels of difficulty, and two levels of prestige source.

## RESULTS

The results of the main analysis are presented in Table 1. Table 2 presents the means

[4] Thanks are due to W. M. Lepley for making a copy of his test available for this research.

TABLE 2
MEAN PROPORTIONS OF SUGGESTIONS ACCEPTED

| Groups | Easy | Difficult | Total |
|---|---|---|---|
| 1. Females: | | | |
| A. Low Prestige: Grade 10 | 37.95 | 59.12 | 48.54 |
| Grade 11 | 24.20 | 46.25 | 35.23 |
| Grade 12 | 18.95 | 33.77 | 26.36 |
| Total | 27.03 | 46.38 | 36.71 |
| B. High Prestige: Grade 10 | 34.78 | 63.88 | 49.33 |
| Grade 11 | 47.15 | 72.45 | 59.80 |
| Grade 12 | 20.27 | 36.08 | 28.18 |
| Total | 34.07 | 57.47 | 45.77 |
| Total All Females | 30.55 | 51.92 | 41.24 |
| 2. Males: | | | |
| A. Low Prestige: Grade 10 | 21.20 | 50.16 | 35.68 |
| Grade 11 | 22.18 | 39.00 | 30.59 |
| Grade 12 | 15.63 | 37.69 | 26.56 |
| Total | 19.67 | 42.28 | 30.98 |
| B. High Prestige: Grade 10 | 33.26 | 55.97 | 44.62 |
| Grade 11 | 39.49 | 57.87 | 48.68 |
| Grade 12 | 10.83 | 19.77 | 15.30 |
| Total | 27.86 | 44.54 | 36.20 |
| Total All Males | 23.76 | 43.41 | 33.59 |
| 3. Grand Total | 27.15 | 47.67 | 37.41 |

of *Ss*' suggestibility scores (proportion of easy and difficult items answered in the way indicated by the incorrect suggestions).

*Suggestibility and Age-Grade Level.* Tables 1 and 2 indicate a significant difference in the acceptance of suggestions as a function of age-grade level. A Tukey gap test applied to these data indicate that Grade 12 *Ss* are significantly different from Grades 10 and 11 ($p < .001$), but these latter two grades do not differ significantly from each other. Thus the significant variance associated with Grade appears to be mainly the result of the relatively low suggestibility of Grade 12.

*Suggestibility and Prestige Level.* Table 1 indicates a significant effect associated with the prestige status of the suggestion source, although confidence in this effect must be lower than it is in the other tabled relationships. In general, there was a tendency for *Ss* to accept suggestions more frequently if the source of the suggestion had high prestige than if the source had low prestige. This was particularly true for the Grade 11 *Ss*, who were largely responsible for the significant Grade × Prestige interaction. While tenth graders accepted very few more suggestions from a high prestige source (about 5% more), and twelfth graders actually were slightly less suggestible when the prestige source was high (they accepted about 7% fewer high prestige than low prestige suggestions), the eleventh grade students were markedly affected by the high prestige, accepting about 20% more high prestige suggestions than low ones. Apparently the prestige of a high school senior has much greater influence on an eleventh grade student than does an eleventh grader on a tenth grader, or a college freshman on a high school senior. These data indicate that where prestige is low, there is a decreasing tendency to accept suggestions as one increases in age-grade status, but that high prestige can alter this trend. This strong interaction of Grade and Prestige masked the differences in suggestibility between Grades 10 and 11; when one applies the Tukey gap test to the suggestibility scores of tenth, eleventh, and twelfth grades in the low prestige condition only, one finds that each grade is significantly different from the others, in the direction of decreasing suggestibility with increasing grade ($p < .001$ in each comparison).

*Suggestibility and Sex.* Tables 1 and 2 indicate that girls accepted significantly more suggestions than did boys. Examination of the Sex × Prestige × Grade interaction reveals that the influence of sex was not constant over all conditions. The difference between boys and girls in suggestibility diminishes with increasing grade level in the low prestige groups, such that the girls accepted significantly more low prestige suggestions than did boys at the tenth grade, with the difference between boys and

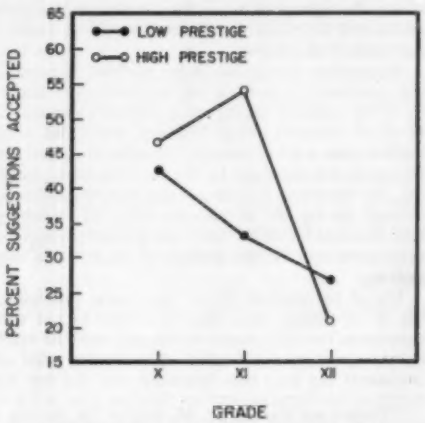


FIG. 1. Interaction of prestige and grade.

girls diminished but still significant at eleventh grade, and the difference between the sexes nonsignificant at twelfth grade. However, in the high prestige groups, girls remain more suggestible than boys throughout the three grades.

*Suggestibility and Difficulty Level.* Tables 1 and 2 reveal a significant difference in the direction of accepting more suggestions on items that *S*s considered to be difficult than on those they thought to be easy. Thus ambiguity of the situations in which *S* must perform is a potent factor in producing dependence on suggestions. Examination of the significant Difficulty × Prestige × Grade interaction indicates that high prestige had a greater effect on the acceptance of suggestions on difficult items in the twelfth grade *S*s than it did on easy items for that grade, whereas in the other two grades, there was no apparent interaction of prestige and difficulty of the item. In general, twelfth graders seemed little influenced by suggestions from either high or low prestige sources, when the items were considered easy. But on difficult items, they did respond differentially to the two prestige sources, so that on difficult items they accepted as many high prestige suggestions as did tenth graders. Thus difficulty and prestige in interaction were able to mask completely the general tendency of twelfth graders to be relatively independent of suggestion.

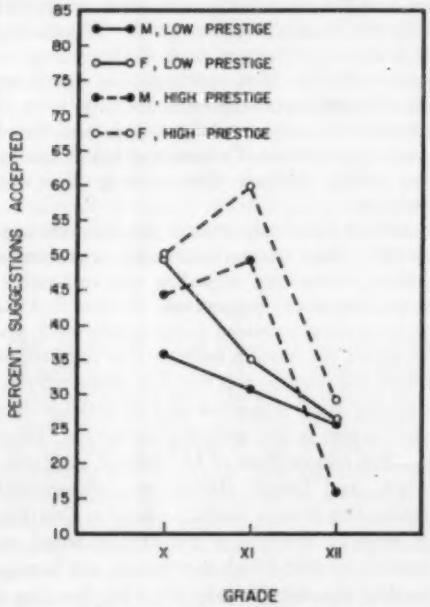


FIG. 2. Interaction of prestige, grade, and sex.

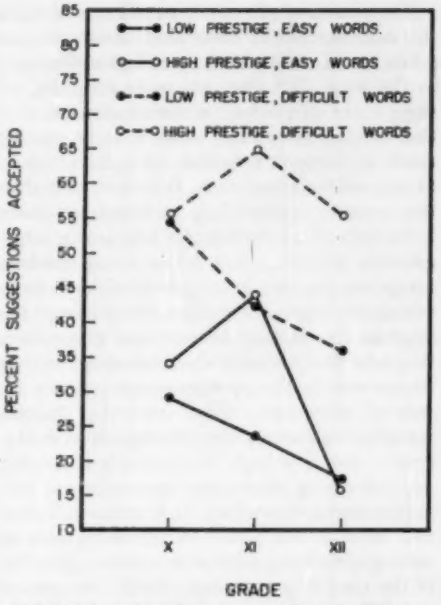


FIG. 3. Interaction of prestige, grade, and difficulty.

## DISCUSSION

In this study, a grand total of 37% of the items on which misleading suggestions were provided were answered in the suggested direction. On roughly two-thirds of the items, even with suggestions provided, *S*s responded without recourse to the "hints." However, the study indicates that this over-all figure of 37% could be inflated or deflated considerably by manipulation of the proper variables, so that among tenth and eleventh grade girls given high prestige suggestions on difficult material, an average of 68% of the suggestions were accepted, and among twelfth grade boys, an average of only 23% of the suggestions made on items considered to be easy were accepted. It thus appears that sex, age-grade status, prestige status of the suggestion source, and difficulty or ambiguity of the problems facing *S*s are powerful variables in affecting *S*'s decision to accept or not accept suggested solutions.

The results of this study are in accord with the commonly-held view that at lower age-grade levels, children are more impressionable, in the sense that they are more accepting of suggestions as guides for their behavior, and that as they grow older there is an increasing tendency for them to become more independent of suggestions from others. It is of interest that this greater suggestibility at lower age-grade levels is particularly true for low prestige suggestions, implying that when young children accept suggestions, it is without reference to or evaluation of the adequacy of the source of the suggestions. As they become less suggestible, they also become more discriminating, so that they come to accept suggestions only in restricted situations: those involving difficult material and where the prestige of the suggestion source is high. Perhaps this is another way of saying that they become more self-reliant and independent. It is of interest that this trend is not based on increasing ease in solving problems, with a concurrent lowering of the need to accept suggestions. In terms of the design of this study, difficulty level of the problems was evaluated independently of age-grade status.

Eleventh graders in this study showed a somewhat peculiar pattern in their responsiveness to suggestions. When the suggestions were ascribed to a low prestige source, the eleventh grade *S*s went along with the general downward trend of suggestibility with increasing age-grade level. However, when the suggestions were ascribed to high school students in their senior year, eleventh graders accepted more suggestions than did any other group. The finding suggests that for eleventh graders, seniors may be much more prestigeful than eleventh graders are for tenth grade students or college freshmen for seniors. This is another way of saying that one year increments in age-grade status are not psychophysically equivalent or equal intervals so far as prestige effects are concerned. It is quite reasonable to suppose that in terms of the social structure of a high school, seniors occupy the highest prestige position, whereas discriminations between sophomores and juniors are not so readily made. As far as the tenth year students are concerned, juniors are not so very different from themselves, whereas the senior year is an outstanding goal for all of the students, including juniors, who are well aware of the differences between themselves and those who are about to be graduated. Further, because seniors do occupy the high prestige position in school, the effects of the status of college freshmen on themselves may be nullified by their own high self-evaluations. The lack of prestige influence by college freshmen may in part also be attributed to the fact that freshmen are perceived even by their educational inferiors as occupying a particularly nonprestigeful position in the college world, in which they are required to wear marks of their low status ("beanies") and are the continual butts of jokes and leg pulling by the more advanced students. These findings suggest that the invocation of "getting into college" as a motivating goal state by high school teachers may be less effective than referring to "being promoted to the senior year."

The increase in acceptance of suggestions as a function of difficulty level of the problems presented to *S*s provides some confirmation of the idea that conformity experiments, in which ambiguity is purposely maintained, inherently increase the probability that *S*s will accept the influence provided by "stooge" *S*s. This finding is in direct agreement with Coffin, who presented his *S*s with mathematical problems after having first categorized his *S*s in terms of amount of mathematical training, and found greater acceptance of erroneous "hints" among the poorly trained than among the well trained.

Girls in this study were in general more suggestible than boys, confirming a common cultural stereotype regarding the responsiveness of females to suggestions. The fact that the girls, as they increased in age-grade level, approached the boys in resistance to suggestions where the prestige was low, but remained consistently more responsive to high prestige than boys is also in line with this stereotype. These findings add to those of McClelland, Atkinson, Clark, and Lowell (1953) that achievement motivation is more readily aroused in girls than in boys by invoking a standard of social acceptability with which to compete, and is more readily aroused this way than by invoking a

"standard of excellence." It may be that in general, American girls are more responsive to social stimulation and status than are boys. However, the greater suggestibility of girls than of boys is not the result of a failure on the part of girls to reduce their responsiveness to high prestige suggestions in twelfth grade. Such a reduction is in evidence, so that twelfth grade girls are only slightly more responsive to high prestige than to low prestige suggestions. However, twelfth grade boys are actually less responsive to high than to low prestige suggestions. The data indicate what may almost be a reaction formation against prestige in the senior boys; they accept more low prestige suggestions than they do high. This reversal, not occurring in the girls, thus makes boys less responsive to high prestige suggestions than girls, although the sexes are equally responsive to low prestige suggestions.

In general these data indicate that both normative and informational aspects of influence are capable of producing modifications in *S*s' behavior. Where *S*'s level of knowledge is low, acceptance of suggestions from both high and low prestige sources is high, and acceptance of suggestions is higher when they come from more prestigeful sources. The data also suggest that although males and females of the same grade level, working on material of equal difficulty, may be considered to have approximately the same amount of knowledge relative to the task, they differ in their acceptance of suggestions. These findings therefore imply that they differ in their responses to the normative aspects of influence, rather than the informational, such that females are more responsive to high prestige than males.

## SUMMARY

An experiment was conducted using boys and girls in Grades 10, 11, and 12 to explore factors associated with the acceptance of suggestions in a situation in which *S*s were given the option of accepting or not accepting the suggestion, and in which task difficulty was varied independently of age-grade status. The variables explored were age-grade status, sex of *S*, difficulty level (or ambiguity) of the task on which *S* was to perform, and prestige level of the suggestion source. Suggestions of high and low prestige were given by penciling in the incorrect answers to some of the items on the Lepley Synonym Vocabulary Test, and ascribing the marked answers to a failure to follow directions to not mark the questions in a class that supposedly had taken the test the previous hour. Prestige level of the suggestions were varied by describing the class which had supposedly taken the test earlier as one year higher or lower in school than the *S*s. Difficulty level was evaluated by having *S*s later rate the words, for which suggested incorrect solutions were provided, as difficult or easy; each *S*'s scores were expressed as proportions of easy and difficult words on which he or she accepted the incorrect suggestions. Results were as follows:

1. Acceptance of suggestions was higher in girls than in boys.

2. High prestige suggestions were accepted more frequently than were those ascribed to a low prestige source.

3. Suggestions were accepted more frequently on difficult than on easy items.

4. Twelfth grade *S*s accepted significantly fewer suggestions than did tenth or eleventh grade *S*s, who did not differ significantly from each other, except when high prestige groups were excluded from analysis. When suggestion source was of low prestige, the three grades differed significantly from each other, in the direction of decreasing acceptance of suggestions.

5. Eleventh grade *S*s accepted many more high prestige suggestions than did tenth or twelfth grade *S*s, suggesting that seniors are more prestigeful to eleventh graders than eleventh graders are to tenth graders, or college freshmen to seniors.

6. The differences between boys and girls in acceptance of suggestions diminished with increasing age-grade status, where the prestige level of the suggestion source was low, but girls remained higher than boys in the acceptance of high prestige suggestions. Boys in the senior year were more resistant to high than to low prestige suggestions, while girls remained more responsive to high prestige.

7. In the twelfth grade, high prestige suggestions had a greater effect on difficult items than it did on easy items. Resistance to high

prestige suggestions could be completely overcome in this group by high item difficulty.

## REFERENCES

COFFIN, T. E. Some conditions of suggestion and suggestibility: A study of certain attitudinal and situational factors influencing the process of suggestion. *Psychol. Monogr.*, 1941, **53**(4, Whole No. 241).

DEUTSCH, M., & GERARD, H. B. A study of normative and informational social influences upon individual judgment. *J. abnorm. soc. Psychol.*, 1955, **51**, 620–636.

LEPLEY, W. M. The rationale, construction, and preliminary try-out of the Synonym Vocabulary Test. *J. Psychol.* 1955, **39**, 215–225.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

MOORE, H. T. The comparative influence of majority and expert opinion. *Amer. J. Psychol.*, 1921, **32**, 16–20.

SAMELSON, F. Conforming behavior under two conditions of conflict in the cognitive field. *J. abnorm. soc. Psychol.*, 1957, **55**, 181–187.

SHERIF, M. *The psychology of social norms.* New York: Harper's, 1936.

(Received September 8, 1959)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 419-423

# THE ATTRIBUTION OF SIMILARITY: THE INFLUENCE OF SUCCESS AND FAILURE[1]

ANTHONY J. SMITH
*University of Kansas*

A new approach to the understanding of interpersonal relations has been provided by Heider (1958) in his conception of balanced states. Several recent investigations provide empirical support for Heider's ideas. They are consistent with the theory in their finding that in the triadic relationship of "p," "o," and "x," referring usually to the subject (p), another person (o), and an act or characteristic (x), respectively, subjects complete a partially defined triad in such a way as to produce cognitive equilibrium or balance (Smith, 1957). The magnitude of the force toward equilibrium has been shown to be a function of the valence for p of the two entities o and x (Smith, 1958). Furthermore, triads that are wholly defined in such a way as to provide balance are reported as being more pleasant than those which according to the theory are unbalanced (Jordan, 1953). In this instance, ratings of pleasantness are taken as crude measures of psychological equilibrium. Finally, group members in reporting their own sociometric choices and their predictions of others' choices yield results that conform to the theory of cognitive balance (Kogan & Tagiuri, 1958).

However, a question might be raised concerning the generality of the tendency toward balance. Are predictions based on the theory of balanced states to be supported in circumstances other than the essentially neutral contexts found in the experiments cited? Furthermore, what effect does an individual's experience of success or failure have on his tendency to attribute similarity to similar and dissimilar persons?

Schachter has reported a series of investigations concerning the need for affiliation (Schachter, 1959) based on earlier work by Festinger (1957). These studies in conjunction with the previously cited research provide a basis for speculating concerning the relationship between the two sets of variables. Schachter found that when persons were given the choice of being alone or being with others they preferred to be in the company of others. Furthermore, when given a choice of companions they tended to select those who were presumably most like themselves (with respect to participation in the experiment). Finally, under conditions of high anxiety induced by threat the need for affiliation with others was increased.

Schachter's first two findings can be related to the results of studies referred to earlier. When p's are asked to indicate their willingness to associate with various o's half of whom are similar, the remainder dissimilar, their average judgment corresponds to mild acceptance. Furthermore, the strength of acceptance is a function of the degree of similarity. Moderate acceptance is accorded similar o's whereas dissimilar o's are responded to with indifference (Smith, 1957).

Schachter's third finding makes it reasonable to assume that p's who experience failure should as a result of the threat involved also experience increased anxiety and have a greater need for affiliation. The increased need should in turn influence the valence of various o's. Festinger's observation that p's who experience moderate dissonance tend to avoid situations that would increase dissonance and seek situations that reduce it indicates that the initial valence of o is crucial. Similar o's who have been demonstrated to have positive valence for p would be expected to have higher valence following p's failure. The effect of failure on p's perception of dissimilar o's is equivocal. Both Heider's and Festinger's theories predict that dissimilar o's should have negative valence. Failure should in turn increase the negative valence. On the other hand, empirical evidence indicates that dissimilar o's are perceived to have neutral valence. In this

[1] This study was supported by a grant from the University of Kansas Research Fund. The author wishes to acknowledge the assistance of Ronald Sobol and John Park.

instance it is impossible to predict whether failure would have no effect on valence or would induce negative valence. Whether one follows the theory or the empirical evidence the differential in valence for similar and dissimilar o's should be greater for the subjects experiencing failure than for subjects experiencing success.

On the basis of these observations and the earlier finding that there is a positive relationship between valence and the attribution of similarity (Smith, 1958), the hypothesis is advanced that those p's who experience failure should show a greater differential in the attribution of similarity to similar and dissimilar o's than should p's who experience success.

It is also hypothesized that an earlier finding derived in a neutral setting to the effect that p should attribute greater similarity to similar o's than to dissimilar o's can be duplicated in a nonneutral setting.

Finally, though no specific hypothesis can be proposed, it is possible to determine whether success and failure have differential systematic effects on the tendency to attribute similarity to similar and dissimilar o's.

## METHOD

### *Subjects*

The subjects in this investigation, 128 undergraduate students enrolled in a course in general psychology, were drawn in equal proportion from each of two summer session classes and an intervening class in the spring semester. They did not volunteer but were required to participate in outside research as an integral part of their introduction to the field of psychology. The Allport-Vernon-Lindzey Study of Values was administered to all class members during the first meetings of the discussion sections of the course as part of their course content.

### *Procedure*

Pairs of unacquainted like-sexed subjects were assembled 6 to 12 weeks after the administration of the test. They were informed that while psychologists had made significant progress in the study of the nonpersonal aspects of intelligence, e.g., numerical ability and spatial visualization, the relationship between such factors and the social expression of intelligence had been only partially explored. It was the purpose of this research to increase our understanding of this relationship. They were to be asked to work on four tasks, the first three of which were standard tests of intelligence taken from the Factored Aptitude Series (King, 1950). Their percentile scores would be revealed to them immediately following each test. It was suggested that their scores might be improved if they were to compete against one another.

Prior to the appearance of the subjects one was designated at random as the successful subject, the other as the failure. Following the administration of the first test (Judgment) the successful subject was told that he had received a percentile rating of 89, while the failure subject was told that his percentile rating was 28. On the second test (Dimension) they received percentile ratings of 85 and 30. The results reported to them for the third test (Memory) were percentile scores of 95 and 16, respectively. It was assumed that these patterns of performances would tend to induce in college students experiences of success and failure. The third task, Memory, was selected to provide a reasonable transition from the typical factored intelligence tests to a more socially oriented task.

Following this the subjects were given the final task. They were reminded that at best they always had limited information concerning other persons. On the basis of this information they were often required to infer other characteristics. The accuracy of such inferences was a measure of one form of social intellectual functioning.

Each subject was then given a partially completed Allport-Vernon-Lindzey test booklet which he was asked to study until he had developed a fairly clear impression of the person from whom the test information had been derived. Then he was requested to complete the test booklet as he believed the other person had. Each person was informed that this "other person" was actually his colleague.

In reality these experimental test booklets were derived from the subject's original test booklet. The partially completed test booklets were prepared so as to provide a systematic degree of similarity or dissimilarity. The method has been described elsewhere (Smith, 1957). Half of the subjects in the successful group were chosen at random to receive similar booklets. The remainder were provided with dissimilar test booklets. A similar procedure was used for the failure group.

In Part I of the test all information on any item involving the subject's highest value was excluded. In Part II for each such relevant item the rank assigned to this crucial value was excluded as was the rank two steps removed. By scoring this crucial scale on the experimental booklet it was possible to determine to what extent the subject attributed to the other person values that were like his own. The absolute difference between the subject's own score and that which he attributed to his colleague was taken as a measure of attributed similarity with smaller values representing greater attribution of similarity.

After these data had been collected the actual nature of the experiment was described to the subjects and each one was given the scores earned in the first three tasks.

## RESULTS

Before examining the hypotheses, three preliminary analyses were undertaken.[2] It was

[2] In each of the analyses to be reported it was neces-

TABLE 1

ANALYSIS OF VARIANCE OF ATTRIBUTION OF SIMILARITY SCORES

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| Success-Failure | 1 | 342.89 | 5.94* |
| Similar-Dissimilar | 1 | 1803.75 | 31.24** |
| Su-Fa × Si-Di | 1 | 135.09 | 2.34 |
| Within | 124 | 57.73 | |
| Total | 127 | | |

* $p < .02$.
** $p < .001$.

TABLE 2

MEANS OF DIFFERENCES BETWEEN OWN AND ATTRIBUTED SCORES FOR SIMILAR AND DISSIMILAR INDIVIDUALS UNDER CONDITIONS OF SUCCESS AND FAILURE

| Conditions | Hypothetical Characteristics of Colleague | | |
|---|---|---|---|
| | Similar | Dissimilar | Total Group |
| Success | 8.38 | 17.94 | 12.84 |
| Failure | 7.17 | 12.61 | 9.94 |
| Total Group | 7.67 | 15.12 | 11.39 |
| Previous Study | 7.05 | 16.18 | 11.61 |

first determined that the original Allport-Vernon-Lindzey test scores on the crucial scales were homogeneous for the four experimental groups ($F = 1.96$; $p > .05$). The differences between experimenters and among replications were then tested in separate factorial analyses of variance. In each case main effects and interaction terms involving main effects were found to be insignificant.[2]

In the final analysis the data were arranged in a 2 × 2 factorial analysis of variance design with success-failure and similar-dissimiiar as the main effects. The results of this analysis as presented in Table 1 reveal that success and failure have significant differential effects on the attribution of similarity ($p < .02$) with greater similarity being attributed by failure subjects. Moreover, the degree of similarity or dissimilarity of the partially completed test booklets is found to be related to a very significant degree to the amount of attributed similarity ($p < .001$) with greater additional similarity being attributed to the presumably similar colleagues. Finally, the interaction effect is found to be insignificant ($p > .15$).

These results are more clearly shown in Table 2 which presents the means for the various groupings. It can be seen that subjects in the failure group attributed greater similarity to their colleagues than did subjects in the success group for both the similar and dissimilar conditions. The differential between similar and dissimilar colleagues is 5.45 for the failure group as compared with 9.56 for the success group.

It is possible to get some crude idea concerning the locus of the effects of success and failure by comparing these results with those obtained earlier under neutral conditions (Smith, 1957). Exact comparisons are not possible because the earlier study used the same subjects for both similar and dissimilar conditions and did not involve the use of intelligence tests or success-failure. Under these neutral conditions (see Table 2) the means of the absolute difference between own and attributed scores for hypothetical similar individuals was found to be 7.05; for dissimilar individuals, 16.18, with a differential of 9.13.

A comparison of the three means obtained for the condition of similarity (8.38, 7.16, and 7.05) reveals substantial agreement. A comparison of the differentials shows very similar results for the success and neutral conditions (9.56 and 9.13) and an appreciably lower value for the failure condition (5.45). The effect of success and failure apparently occurs primarily in the failure condition as a result of the attribution of greater similarity to the dissimilar individuals (12.61) than occurs in the success condition (17.94). The difference between these two means is found to be significant ($t = 2.50$; $p < .02$). The success condition in turn is found to resemble the neutral condition (16.18).

## DISCUSSION

In terms of Heider's theory, a subject who examines a similar test booklet can experience for each item the Condition pLx, oLx or p ~ Lx, o ~ Lx. In each instance there is a

sary to determine whether the assumption of homogeneity of variances could be met. This was tested by means of Hartley's procedure (Walker & Lev, 1953). In each instance the hypothesis of homogeneous variances was found to be acceptable.

[2] Tables summarizing these results are omitted in order to conserve space. The unreported results are essentially duplicated in the analysis of variance reported in Table 1.

tendency to attain balance in the pox triad. This is accomplished when pLo. Thus there is a cumulative effect in inducing a force corresponding to pLo. When a subject is required to predict his colleague's response, two elements are known: his own response to x and pLo. Balance is achieved when he attributes his own response to o (pLx, oLx or p $\sim$ Lx, o $\sim$ Lx). When, on the other hand, he examines a dissimilar test booklet, he observes for each item either pLx, o $\sim$ Lx or p $\sim$ Lx, oLx. Balance occurs when p $\sim$ Lo. When he estimates unknown characteristics of o, he can achieve balance by attributing to o characteristics other than his own (pLx, o $\sim$ Lx or p $\sim$ Lx, oLx). It should be noted, however, that if the lower rankings or weightings on the Allport-Vernon-Lindzey instrument represent indifference rather than dislike, it can only be predicted that the acceptance of dissimilar o's should be less than the acceptance of similar o's.

The first hypothesis to be examined in the present study states that the differential in attribution of similarity to similar and dissimilar individuals should be greater for subjects experiencing failure than for those experiencing success. An examination of the interaction term that reflects this relationship (Success-Failure $\times$ Similar-Dissimilar) gives no support to the hypothesis. In fact, the trend is in the opposite direction.

It should be recalled that this hypothesis was based on the assumption that the difference between the valences of similar and dissimilar persons would be increased following the experience of failure. The hypothesis further requires that the initial valence of the dissimilar person be neutral or negative.

The experiment was originally designed with the expectation that similar and dissimilar test booklets would completely define the degree of similarity. In retrospect, however, it can be seen that before any experimentally introduced evidence was available concerning similarity, each subject could easily observe that he and his colleague were either alike or indistinguishable with respect to many characteristics, e.g., sex, age, race, educational status. In each instance in which he perceived a shared characteristic (pUx, oUx), balance would be attained when the Condition pLo was established. Therefore following the introduction of dissimilar test booklets which ordinarily generate neutral valence, one would now expect a low degree of positive valence. If this is the case, failure should result in increased positive valence for the dissimilar subjects and no significant interaction between success-failure and the attribution of similarity would be expected.

From this line of reasoning one would also infer that the attribution of similarity should be greater for subjects experiencing failure than for those experiencing success. An examination of the significance of the main effect corresponding to success-failure indicates that the results are consistent with this inference. Moreover, the difference occurs for both similar and dissimilar individuals though only in the latter case is it significant.

Turning to the second hypothesis, the results indicate that in responding to hypothetical persons subjects attribute significantly greater added similarity to similar individuals than to dissimilar individuals. The $t$ of 5.59 obtained in the present study compares with a similar finding ($t$ = 6.12) reported in an earlier study (Smith, 1958). This finding provides empirical support for the hypothesis that predictions based on the theory of balanced states are confirmed not only in neutral contexts but in the nonneutral contexts corresponding to success and failure.

## SUMMARY

Sixty-four pairs of unacquainted like-sexed subjects were given a series of three tests of intellectual ability. One subject in each pair was informed that he had earned high percentile scores on each task, while the other subject was told he had earned low scores. Each subject then received a partially completed Allport-Vernon-Lindzey test booklet designed either to be identical with or systematically dissimilar to the one completed by the subject 6 or more weeks earlier. *S* was asked to study this test booklet presumably derived from his colleague until a relatively clear conception of the person emerged. Then he was to complete the booklet as he believed his colleague had. Degree of similarity attributed was defined as the absolute difference between own and attributed score on a crucial scale.

The results revealed that failure subjects at-

tributed greater similarity to their colleagues than did successful subjects, particularly with respect to presumably dissimilar colleagues.

A previous finding in a neutral context involving the greater attribution of similarity to similar persons than to dissimilar persons was confirmed in the present nonneutral context.

The predictions that persons experiencing failure will show a greater differential in the attribution of similarity to similar and dissimilar individuals than will persons experiencing success was not supported. A possible explanation was offered.

## REFERENCES

FESTINGER, L. *A theory of cognitive dissonance.* Evanston: Row, Peterson, 1957.

HEIDER, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

JORDAN, N. Behavioral forces that are a function of attitudes and of cognitive organization. *Hum. Relat.*, 1953, **6**, 273–287.

KING, J. E. *Factored Aptitude Series.* Chicago: Industrial Psychology, Inc., 1950.

KOGAN, N., & TAGIURI, R. Interpersonal preference and cognitive organization. *J. abnorm. soc. Psychol.*, 1958, **56**, 113–116.

SCHACHTER, S. *The psychology of affiliation.* Stanford: Stanford Univer. Press, 1959.

SMITH, A. J. Similarity of values and its relation to acceptance and the projection of similarity. *J. Psychol.*, 1957, **43**, 251–260.

SMITH, A. J. Perceived similarity and the projection of similarity: The influence of valence. *J. abnorm. soc. Psychol.*, 1958, **57**, 376–379.

WALKER, HELEN M., & LEV, J. *Statistical inference.* New York: Holt, 1953.

(Received September 8, 1959)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 424-431

# THE INDIRECT REPRESENTATION OF SEXUAL STIMULI BY SCHIZOPHRENIC AND NORMAL SUBJECTS[1]

LEO B. AREY

*Veterans Administration Hospital, Hines, Illinois*

THIS study is concerned with the nature of the perceptual hypotheses advanced by schizophrenics and normals when confronted with sexual pictures presented at a speed just below the time threshold for conscious recognition and report. The method of tachistoscopic experimental attack was suggested, in part, by the work of Poetzl (1917), whose findings have been substantially confirmed by the recent studies of Shevrin and Luborsky (1958) and Fisher (1954, 1956). Poetzl's results indicated that stimuli apparently perceived outside of the focus of conscious awareness could later appear, symbolically transformed, in the manifest content of his subjects' dreams. One of the intriguing problems suggested by the Poetzl data was the temporal locus of these symbolic transformations. In the course of a preliminary experiment, directed toward the testing of a hypothesis of levels in symbolism, and involving the tachistoscopic showing of sexual pictures to schizophrenic and normal subjects, I observed an interesting phenomenon. In certain threatening areas the person responded to inadequately perceived sexual stimuli with perceptual distortions couched in images that were often remarkably similar to dream symbols described in the psychoanalytic literature. I observed this phenomenon many times before it finally occurred to me that I might be witnessing an immediate symbolic transformation of perception. Further, while these hypothetical symbolic responses of normal and schizophrenic subjects were often alike in quality, there were sometimes striking differences in the reported imagery.[2]

My observations led to the formulation of two general hypotheses that constitute the conceptual background of the current experiment:

1. When a person is confronted tachistoscopically with pictures representing various aspects of sexuality, the antecedents of objectively accurate conscious perception are often images that indirectly represent the stimulus object or activity.

2. These images reflect significant aspects of the subject's fantasies and affective attitudes concerning the stimulus object or activity, i.e., the kind of image that is used suggests part of the unconscious meaning the indirectly represented sexual object or activity has for the individual. If these images are considered as models of the way in which the subject unconsciously perceives sexual objects, it is then possible to judge whether his more active fantasies concerning these objects are such as to sustain and facilitate or to inhibit and obstruct his attempts to achieve satisfying interpersonal relationships and to maintain a relatively adequate integration of personality. An image can be rated as "sustaining" (healthy) if it reflects a fantasy of the sexual object or activity as a source of libidinal gratification and support; it can be rated as "nonsustaining" (pathological) if it reflects a fantasy of the sexual object or activity as destructive, threatening, or non-

[1] This paper is based on a doctoral dissertation submitted to the University of Chicago in 1952. The experimental work was carried out while the author was a psychology trainee at the Veterans Administration Hospital, Hines, Illinois. Reviewed in the Veterans Administration and published with the approval of the Chief Medical Director. The statements and conclusions published by the author are the result of his own study and do not necessarily reflect the opinion or policy of the Veterans Administration.

Morris I. Stein, James G. Miller, Roy Brener, and David Shakow have all contributed invaluable advice, criticism, and moral support for which the author is deeply indebted. The writer is also grateful to Donald W. Fiske, who made many valuable suggestions concerning the statistical treatment of the results of this experiment.

Finally, this study owes much to the Psychiatry Service at the Veterans Administration Hospital, Hines, Illinois, for the cooperation of its staff in the selection of experimental subjects, and to the clinical psychology staff members and trainees at the Chicago Veterans Administration Mental Hygiene Clinic who served as judges in this experiment.

[2] This study is unpublished, but its design and results are described in Ch. I–II of my doctoral dissertation (Arey, 1952).

gratifying. The sustaining image will suggest that the subject's response to the object is essentially adient, with positively toned affective attitudes. The nonsustaining image will suggest that the subject's response to the object is predominantly abient, with fearful or negatively toned attitudes; or adient, with hostile or fearful attitudes. The characteristics that the subject attributes to the object may be projections of his own impulses toward the object, but the point is that these characteristics, whether they belong to the subject or the object, should somehow be expressed in the image used to represent the object.

The specific experimental hypothesis, in its most conservative form, reads: assuming that schizophrenics are more threatened by various aspects of sexuality than are normals, and assuming that the prerecognition distortions of tachistoscopically presented sexual stimuli are often indirect representations reflecting the subject's immediate affective response to the stimulus object, then a group of normal subjects can be differentiated from a group of schizophrenic subjects on the basis of the affective attitudes implied in the imagery of their respective prerecognition distortions.

The experiment was originally planned to deal only with the perceptual distortions that could be reasonably judged as "true symbols," but it became evident that it was going to be extremely difficult to assign limits to what constituted a "true symbol." In many instances it was impossible to differentiate between "true symbolism" and other forms of indirect representation, such as substitution of the part for the whole, allusion, plastic imagery, preconscious or conscious metaphor, etc.[3] Yet, excluding any of these distortions would have left too much room for the operation of unconscious biases in the selection of data. Consequently, both the experimental and the two general hypotheses, which were originally phrased in terms of "symbols" and "symbolism," are now stated in terms of "imagery" and "indirect representation." Although the judges were instructed to treat *all* the distortions as symbols, it is clear that they were dealing with the general class of indirect representations, and the interpretation of the experimental data must take this into account.

Another problem was posed by the distortions that might be mere misidentifications, with no symbolic or indirect representational reference. Some of these nonrepresentational responses to sexual stimuli were almost surely included in the data, but similar nonrepresentational responses to nonsexual stimuli provided a control for their effects.

## METHOD

### *Design*

The hypotheses were tested by using an experimental group of 25 schizophrenic male subjects and a control group of 25 normal male subjects, with the groups equated for age, intelligence, ethnic group, and education. Each subject was shown a series of 24 pictures representing various aspects of sexuality (sexual intercourse, birth, the genitalia, etc.) and he was asked to describe what he saw at .01-second presentations of the stimuli. As a control against the possibility that the responses given to these stimuli were not a specific function of the nature of the ideational areas represented, 9 blank cards and 18 nonsexual pictures were also presented.

### *Subjects*

*The experimental group* consisted of 25 male schizophrenic subjects from the psychiatric wards at the Veterans Administration Hospital, Hines, Illinois. Criteria for inclusion were:

1. The psychiatrist in charge of the patient must express reasonable certainty that in his opinion the patient can be classified as schizophrenic, and another psychiatrist familiar with the case must confirm this diagnosis.

2. The clinical record must not reveal any suspicion of organic trauma or of mental defect per se.

3. The patient must be in sufficient contact to communicate with the experimenter and he must demonstrate during the presentation of the practice stimuli that he is cooperative enough to make an attempt to see the stimuli and to report on what he sees.

This last specification is quite important, for much of the psychological research on schizophrenia ends in ambiguity because the researcher often confuses a subject's refusal to participate in a task with an inability to perform it (cf. Bleuler, 1950, p. 58; Cameron, 1944, p. 903). Various internal controls indicate that a high level of cooperation was obtained from the subjects, who were mostly acute treatment cases rather than chronic, long hospitalized patients.[4] Further, no

[3] For an analysis of the criteria differentiating "true symbolism" from other forms of indirect representation, see Jones (1949).

[4] A check of the hospital records approximately 8 months after the collection of the experimental data revealed that seven of the group had been discharged with "maximum hospital benefits," seven were on extended trial visits, one had been discharged "against

TABLE 1
EQUATION OF EXPERIMENTAL AND CONTROL GROUPS

| | Age in Years | | | IQ (as Measured by VIBS Scale) | | | Education (in Years of School Completed) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Range | Mean | $\sigma$ | Range | Mean | $\sigma$ | Range | Mean | $\sigma$ |
| Normal ($N$ = 25) | 21–47 | 31.32 | 8.40 | 83–134 | 111.68 | 14.57 | 7–20 | 11.96 | 3.62 |
| Schizophrenic ($N$ = 25) | 22–52 | 31.36 | 7.97 | 89–137 | 110.44 | 12.52 | 5–15 | 10.62 | 2.43 |
| | $CR$ = .017<br>$p$ = .98 | | | $CR$ = .319<br>$p$ = .75 | | | $CR$ = 1.50<br>$p$ = .13 | | |

patients were included who were currently receiving either electroshock or chemoshock therapy. It required an original starting group of 36 to provide the 25 subjects meeting the criteria for inclusion.

*The control group* consisted of 25 "normal" male subjects. None of these subjects was judged by the experimenter to be obviously severely neurotic or psychotic on the basis of a brief "neurotic inventory" (Clark revision of the Thurstone Personality Schedule) consisting of 25 items, supplemented by observations made while an abbreviated form of the Wechsler-Bellevue Intelligence Scale (VIBS) was being administered. However, there were probably several neurotic individuals and at least one latent schizophrenic included in the control group. These subjects were retained in the group because these opinions concerning them were formed on the basis of performance in the experiment itself, plus material confided to the experimenter after completion of the experimental task.

The groups were equated by first choosing the experimental group, and then selecting the control group so as to be equated with it in terms of age, intelligence, ethnic group, and education. Forty-one normal subjects were tested; five of them were excluded because the experimenter elicited enough evidence of neurotic traits to make the decision to exclude them before the experimental task was begun. In choosing the 25 control subjects from the remaining pool of 36 candidates, the experimenter tried to match the groups as closely as possible in terms of age and IQ (as measured by the VIBS scale). The equations for age and intelligence were quite satisfactory; that for education was less so, but the difference between the groups was not statistically significant. There were 20 white and 5 Negro subjects in each group.

### *Stimuli*

First, there were 10 practice figures, presented for the purpose of acquainting the subject with the use of the tachistoscope and assuring the experimenter that the subject was likely to try to cooperate. These figures were simple designs, e.g., plus sign, circle, triangle, square, arrow.

The test and control stimuli, described in order of presentation to each subject, were:

1. A woman's lips, smiling, with teeth showing[5]
2. Blank card
3. Anatomic drawing, showing a fetus in the mother's womb
4. Blank card
5. A young woman's breasts
6. Blank card
7. The painting by Barend van Orley, "Neptune and the Nymph," an unmistakable representation of two nude figures in heterosexual intercourse
8. Blank card
9. A man's lips, open slightly
10. A cedar chest
11. Rodin's "The Thinker"
12. A little boy, standing, dressed in sleeping suit
13. A doctor, holding a newborn baby upside down
14. A modern house, with wide balcony and straight, overhanging eaves
15. A wood carving, showing a young man masturbating. The figure is shown from the chest to about the middle of the thighs.
16. A formal garden, with paths leading to a central pool
17. A view of Rodin's "The Kiss," chosen to give an impression of heterosexual intercourse
18. A woman smiling, playing with and partly holding a happy looking baby, who is sitting on a bed
19. The male genitalia. The penis is erect.
20. A dining room table. The table is bare except for a bowl of flowers in the center.
21. A woman lying on her back, with legs outspread and with genitalia clearly shown

medical advice," three had been discharged as "absent without leave," three had been transferred to other hospitals, and four of the group were still in the same hospital. Such statistics are difficult to interpret, but they do tell us that 18 of the 25 experimental subjects could be temporarily absorbed at some level by the society outside the hospital within a relatively short time.

[5] Three pictures of an "open mouth" (No. 1, 9, 45) were included in the series for exploratory purposes only, and responses to these stimuli were excluded from the experimental data. No. 45 was a duplicate photograph of the No. 1 stimulus. With this one exception, all the pictures were different despite the identity of some descriptions. Unless otherwise specified, e.g., as a drawing or a painting, the figures were photographs of the subject matter described.

22. The heads of two dogs, one above the other
23. A drawing of the male genitalia. The penis is erect.
24. A chair next to a small table on which there is a desk lamp
25. A young woman's breasts
26. The face of a young, smiling child
27. The nude torso of a young woman, lying down
28. Blank card
29. A painting of Hercules, showing a very muscular man with arms drawn back ready to strike at something or someone
30. A young colt
31. A woman's body, from the chest down. Her legs are spread open, with the genitalia clearly shown.
32. Drawing of three interlocking circles, with the interiors of the circles shaded by pencil markings
33. A view of Rodin's "The Kiss." This can be seen as either a kiss between two nude people or it can be seen as heterosexual intercourse.
34. A small white country church with a black roof, topped by a black steeple supporting a white cross
35. Drawing of the male genitalia. The penis is flaccid.
36. Interior corner of a living room, showing couch, chair, lamp, etc.
37. A young woman's breasts
38. Approximately a dozen books, lined up side by side in a vertical position
39. A statue of a nude youth holding a small, elongated jar just above and in front of the genital area. He is pouring water into his cupped left hand.
40. The face of a beautiful young woman with flowing hair
41. Nude torso of a young woman, with quite prominent breasts
42. A young man sitting on a couch; he has turned to face a young woman standing behind the couch. The woman is offering him a dish of fruit.
43. A Pompeiian bronze; the statue is called "Die Potenz." The head is bent, slightly forward, as the man gazes at his erect penis.
44. A panoramic view of rolling slopes and softly contoured hills
45. A woman's lips, smiling, with teeth showing
46. Photograph of a nest of suitcases. A large suitcase is opened; in it is a smaller suitcase, opened; in it, in turn, another even smaller suitcase is opened. Altogether, there are six of the suitcases.
47. A nude woman, from her crotch to her feet. She is sitting on the edge of a bed, with her legs spread wide.
48. Blank card
49. Drawing of a nude male figure from chest to middle of thighs. His hand grasps his erect penis.
50. Blank card
51. Close-up of the female genitalia. A small doll's head and body protrude from the vaginal opening. At fast tachistoscopic speeds, this picture provides an excellent representation of birth.
52. Blank card
53. Photograph of a nude female torso, with quite prominent breasts
54. Blank card

The ordinal number of presentation was even for all of the control stimuli and odd for all of the test stimuli. These pictures were photographed so that one dimension was always approximately 7 inches; the other dimension, which could be either height or width, was allowed to vary from 3.75 to 7 inches. They were mounted on 10.5 × 8.5 inch white carboard and could be inserted directly into the tachistoscopic viewing field. The nine blank controls were simply white cards with no pictures mounted on them.

Strictly speaking, all of the 18 nonsexual pictures have a potential symbolic sexual reference. But that sexual reference is certainly not the direct and straightforward one of the 24 frankly sexual stimuli.

### *Apparatus and Procedures*

A Harvard (modified Dodge) Tachistoscope was used in the experiment. The exposure field was approximately 23 inches from the subject's eyes. The preexposure field and the exposure field were each illuminated by two white, 4-watt fluorescent bulbs, giving a brightness of 2.8 apparent foot-candles, as measured at the eyepiece of the machine. This illumination was then decreased by four sheets of 20-pound white bond paper placed over each of the four milk screens.[6]

Each subject was tested individually in a small room, with only the experimenter and the subject present. Each subject was instructed:

> Tell me exactly what you see when the picture flashes on in the machine. Even if you are not sure what the picture actually shows, I want you to tell me what your first impression of it is. It is very important that you tell me your impression every time a picture flashes in the machine.

The exposure time for every tachistoscopic presentation was always .01 second. Some of the stimuli were presented several times to subjects in both groups when the experimenter felt that especially significant material could be elicited. Also, some of the pictures correctly recognized by the subject on the first presentation were occasionally re-presented several times so that the subject would not believe that repeated presentation meant that his first response was wrong. All responses were recorded as nearly verbatim as possible. This was fairly easy to do, as most responses were rather brief. Only the responses to the first presentation for each stimulus were considered in the statistical treatment of the data. After the completion of the presentations for each of the first 15 cards, it was shown at unlimited time and the subject was asked for a description of it. The purpose of this procedure was to convince the subject that he need have no hesitation in reporting his perception of a sexually toned stimulus picture when he consciously recognized it as such.

At the conclusion of the experimental task, each

[6] Recognition of tachistoscopically presented stimuli depends on a complex interaction of speed of presentation, brightness of the stimulus field, and the structural qualities of the stimulus. My experience indicates that the phenomenon of indirect representation is best elicited when about 10–20% of the subjects achieved accurate recognition of the stimulus, thus assuring that partial cues concerning the nature of the stimulus are likely to be accessible to the subjects who do not consciously recognize the picture.

subject was given the opportunity to see all the pictures in the series at unlimited time. Every subject in both groups took advantage of this opportunity. Each subject was then given the chance to talk about the pictures and to verbalize some of the feelings that the experimental task had generated.

### *Treatment of Data*

Six independent male judges (consisting of four members of the clinical psychology staff and two advanced VA clinical psychology trainees, all at the Chicago VA Mental Hygiene Clinic) rated the responses given to several of the sexual stimuli. Four judges rated the responses to three of the pictures; one judge rated the responses to four of the sexual pictures; the sixth judge rated the responses to five of the sexual pictures. The experimenter's own ratings of the responses as to imagery level clearly indicated that it was safe to assume that if the judges could differentiate the two groups on the basis of the responses to these few sexual pictures, then they should be able to make differentiations at as high, or higher, levels of confidence by using the responses to the full set of 24 sexual stimuli.

These judges had only the actual responses, with no way of determining whether a response was given by a schizophrenic or a normal except by a judgment of the response itself. The general instructions to the judges were as follows:

> These are the responses given by 25 normals and 25 schizophrenics to this picture when it was presented to them tachistoscopically at .01 second. [The specific picture was then placed before the judge.] There are some recognitions, and there are some gross perceptual distortions and misidentifications of the stimulus picture. The recognitions are to be grouped in one category. Even if you can tell from language cues, or what have you, that the response was given by a schizophrenic, if the response is essential recognition of the central concept of the picture, then it goes into the recognition category. Assume that the perceptual distortions are symbols representing the way the subject unconsciously perceives the object, or the way in which he unconsciously tends to respond affectively to the object. Assume that they are the models that he unconsciously lives by. Assuming this, try to decide whether the subject's impulses toward the object are libidinally or destructively cathected. If the symbol is a libidinal one, indicating that the object is perceived as a source of gratification, then rate the response as "sustaining." By this we mean that this kind of model would tend to sustain the person's attempts at satisfactory interpersonal relationships, as well as tend to sustain a relatively satisfactory level of personality integration. If the symbol is a destructive-fearful one, indicating that the object is feared, hated, or perceived as essentially nongratifying, then rate it as "nonsustaining." By this we mean that this sort of model would tend to interfere with or obstruct attempts at satisfactory interpersonal relationships and the maintenance of a relatively satisfactory level of personality integration.

Only these very general instructions were given to judges for rating Pictures 7, 17, 33, 47, and 51. For Pictures 15 and 43, more specific instructions were given.

For 15 the additional instructions were as follows:

> To be rated as sustaining, the symbol must have some formal similarity to the idea or object symbolized. It must suggest an intact penis. To be rated as nonsustaining, the symbol may express either the idea of castration (as a consequence of masturbation), or it may express an idea of violence beyond the control of the subject, i.e., of forces beyond the subject's control. Also include here all feminine identifications, or symbols suggesting feminine identification of the stimulus (assuming that feminine identification implies castration).

For Picture 43, "Die Potenz," the additional instructions were as follows:

> Assume that the symbol reflects the subject's unconscious phantasy concerning his potency, his unconscious acceptance or nonacceptance of potent masculinity. Will such a phantasy sustain or will it obstruct his attempts at satisfactory interpersonal relationships? Rate the perceptual distortions accordingly.

After all the ratings had been completed and recorded, the two respective imagery scores accruing to each subject on the basis of the judge's ratings were tabulated. Since the hypothesis demands that the two groups of subjects be differentiated by the respective kinds of indirect representation they use for sexual objects and activities, the responses showing adequate recognition and the responses of stimulus rejection (e.g., "Don't know what it is," or "My eyes blinked then," etc.) obviously could not be included in the tabulation of imagery scores. The resulting quantitative difference in the scorable distortions produced by the two groups introduced a source of error in interpreting the judgments. For instance, suppose that in rating responses to five pictures, a judge found that 110 responses from the schizophrenic group were scorable either as sustaining or as nonsustaining images, but that only 100 responses from the normal group were scorable images, with the remainder being rated as recognition or stimulus rejection. Chance factors alone would then favor the schizophrenics' obtaining higher scores than the normal subjects in both categories of imagery. To take care of this difficulty, a specific "correction constant" was derived for each judge, thus: Let $x$ equal the sum of the mean of the sustaining imagery scores and the mean of the nonsustaining imagery scores assigned by the judge to the schizophrenic group on the basis of the responses to the pictures rated by him. Let $y$ equal the sum of these two means for the normal group. Then, the constant, $c$, is found by solving the equation: $cy = x$. The means of the normal group's respective scores for sustaining imagery and for nonsustaining imagery were then multiplied by the obtained constant. It was assumed that the standard deviations of the two distributions were not changed. This device permitted a direct comparison between the means of the schizophrenic group and the normal group for each of the two kinds of imagery scores.

TABLE 2

DIFFERENTIATIONS BETWEEN 25 SCHIZOPHRENICS AND 25 NORMALS MADE BY SIX JUDGES IN ASSIGNING NONSUSTAINING IMAGERY SCORES

| Judge | Code numbers of respective pictures for which each judge rated the responses | Means of nonsustaining imagery scores assigned by each judge to the two groups | | $t$ |
|---|---|---|---|---|
| | | Normals | Schizophrenics | |
| F | 15, 33, 43 | 1.01 (0.82)[a] | 2.08 (0.74) | 4.75*** |
| D | 7, 15, 33, 43 | 1.76 (0.99) | 2.84 (1.08) | 3.62*** |
| E | 7, 15, 43 | .66 (0.63) | 1.40 (0.80) | 3.54*** |
| A | 7, 15, 43 | 1.45 (0.78) | 1.92 (0.93) | 1.88 |
| B | 15, 33, 43, 47, 51 | 1.80 (1.02) | 2.24 (1.14) | 1.42 |
| C | 7, 17, 33 | 1.77 (0.75) | 1.84 (0.88) | 0.30 |

Note.—Judges F, E, A, C rated responses to three pictures. Thus in the rating of these judges, the highest possible score an individual can accrue is 3. In Judge D's ratings of four pictures, 4 is the ceiling for an individual's score; in Judge B's ratings, 5 is the ceiling.

Mean score of normal group is corrected by a specific constant for each judge in order to rule out the effect of the schizophrenics' production of more scorable distortions as a factor in the judges' discriminations between the two groups (see text). The uncorrected normal means: Judge F, .96; Judge D, 1.76; Judge E, .64; Judge A, 1.28; Judge B, 1.64; Judge C, 1.40.

[a] Standard deviations in parentheses.

*** $p < .001$, two-tailed test.

TABLE 3

DIFFERENTIATIONS BETWEEN 25 SCHIZOPHRENICS AND 25 NORMALS MADE BY SIX JUDGES IN ASSIGNING SUSTAINING IMAGERY SCORES

| Judge | Code numbers of respective pictures for which each judge rated the responses | Means of sustaining imagery scores assigned by each judge to the two groups | | $t$ |
|---|---|---|---|---|
| | | Normals | Schizophrenics | |
| F | 15, 33, 43 | 1.55 (0.89)[a] | .48 (0.64) | 4.78*** |
| D | 7, 15, 43, 33 | 1.76 (0.95) | .68 (0.84) | 4.18*** |
| E | 7, 15, 43 | 2.10 (0.87) | 1.36 (0.79) | 3.07** |
| A | 7, 15, 43 | 1.27 (0.71) | .80 (0.80) | 2.15* |
| B | 15, 33, 43, 47, 51 | 2.60 (0.98) | 2.16 (1.22) | 1.37 |
| C | 7, 17, 33 | .53 (0.70) | .48 (0.49) | 0.41 |

Note.—Same as note to Table 2, except that the uncorrected normal means here are: Judge F, 1.48; Judge D, 1.76; Judge E, 2.04; Judge A, 1.12; Judge B, 2.36; Judge C, .44.

[a] Standard deviations in parentheses.

* $p < .05$, two-tailed test.

** $p < .01$, two-tailed test.

*** $p < .001$, two-tailed test.

## RESULTS AND DISCUSSION

Schizophrenic subjects were expected to give more nonsustaining images for sexual objects and activities than would normal subjects, and, conversely, normal subjects should use more sustaining images than should schizophrenics. The results shown in Table 2 and Table 3 confirm the prediction. All six judges assigned more of the schizophrenics' than of the normals' distortions to the nonsustaining category; conversely, all six judges assigned more of the normals' than of the schizophrenics' distortions to the sustaining category.[7] The differentiation between groups was in the predicted direction on both measures for all six judges, and the differentiation between groups was significant beyond the .01 level on both measures for three of the judges.

Two of the judges who differentiated between schizophrenics and normals at high levels of significance in rating the perceptual distortions to the sexual stimuli were asked to rate the responses to a large sample (11) of the nonsexual stimuli. They were unable to differentiate between the two groups at

[7] For Judge C, this is true for sustaining imagery scores only when the Normal mean is corrected to rule out the effect of the schizophrenics' production of more scorable distortions than are given by the normals.

As a background against which to evaluate the ratings of the six independent judges, my ratings of distortions given to the full set of 24 sexual stimuli yielded these differentiations in the predicted directions between normals and schizophrenics: on sustaining imagery scores, $t = 5.78$, $p < .001$; on nonsustaining imagery scores, $t = 5.26$, $p < .001$.

In rating distortions to Pictures 15, 33, and 43 (i.e., those rated by Judge F), my differentiations between groups were: on sustaining imagery scores, $= 5.86$, $p < .001$; on nonsustaining imagery scores, $= 3.90$, $p < .001$.

In rating distortions to Pictures 7, 17, and 33 (i.e., those rated by Judge C), my differentiations between groups were: on sustaining imagery scores, $t = 2.66$, $p < .02$; on nonsustaining imagery scores $t = 1.92$, $p < .1$. Judge C was given his particular group of pictures (all representations of heterosexual intercourse) to see if he could pick up any differences I might have missed in my ratings.

Thus Judge C was given a more difficult task than Judge F, for he was asked to make ratings on the responses to a group of pictures that in my own ratings yielded questionable, borderline differentiation. Judge F, on the other hand, rated responses to pictures that in my ratings yielded differentiations roughly equivalent to those obtained in rating the whole series of 24 sexual stimuli.

Judges D, E, A, and B also were given groups of pictures to rate for which my respective differentiations were roughly equivalent to those obtained in rating the whole series.

statistically significant levels on the basis of these ratings. For one judge, $.10 < p < .20$; for the second judge, $.30 < p < .40$.

There were no differences between the two groups in response to the blank cards. No subject in either group reported seeing anything except a "blank," or a "flash," or a "light" when these cards were presented. One schizophrenic subject did respond "just a flash" to one card and then go on to tell a long story about Flash Gordon, but this narrative was clearly not intended as the description of a perception.

One of the assumptions of the experimental hypothesis is that most people in this culture are threatened by some aspects of sexuality, but that schizophrenics are more threatened by various aspects of sexuality than are normal persons. This assumption was supported by the recognition scores made by the two groups of subjects for the two classes of pictures.

The sexual and the nonsexual stimuli were rated for structural difficulty by two husband-and-wife teams, and it was possible to use their ratings to pick out 13 of the sexual pictures whose structural difficulty was considered as roughly equivalent to that of 13 of the nonsexual stimuli. All responses to both sexual and nonsexual stimuli were rated for recognition versus nonrecognition by the experimenter on the basis of criteria established before any of the data were collected.[8] On these two groups of equated pictures, the normal group recognized significantly more of the sexual pictures than did the schizophrenic group ($t$, 3.63; $df$, 48; $p < .001$). Normal subjects recognized slightly more of the nonsexual stimuli than did the schizophrenics, but the difference was not significant ($t$, 1.33; $df$, 48; $p < .20$). Further, both normal and schizophrenic subjects recognized significantly more of the nonsexual stimuli than of the sexual stimuli ($p < .001$). Finally, the differences in recognition scores for the two classes of pictures were significantly greater ($p < .02$) for the schizophrenic group (mean difference, 4.56) than for the normal group (mean difference, 2.96).

The differences in stimulus rejections by the two groups for the two classes of pictures parallel the differences in recognition. Schizophrenics exceeded normals in stimulus rejections of the sexual pictures ($p < .01$); in fact, none of the normals gave more than one stimulus rejection, while 14 of the schizophrenics gave two or more. Further, schizophrenics gave significantly more stimulus rejections in response to sexual stimuli than to the nonsexual stimuli ($p < .001$).

The differences in recognition scores of the two groups for the two classes of pictures support the assumption that male schizophrenics are more threatened by various aspects of sexuality than are normal young men.[9] While it cannot be experimentally demonstrated that reported recognition accurately corresponds to conscious recognition, it nevertheless seems reasonable to assume that differentials with respect to *report* of recognition are valid measures of differences in the experience and handling of sexual threat.

## SUMMARY AND CONCLUSIONS

On the assumption that male schizophrenics are more threatened by various aspects of sexuality than are normal males, and that the prerecognition distortions of tachistoscopically presented sexual stimuli are often indirect representations reflecting the subject's immediate affective response to the stimulus object, it was hypothesized that a group of normal subjects could be differentiated from a group of schizophrenic subjects on the basis of the *affective attitudes* implied in the imagery of their respective prerecognition distortions.

This hypothesis was tested by using an experimental group of 25 schizophrenic male subjects and a control group of 25 normal male subjects, with the groups equated for age, intelligence, ethnic group, and education.

[8] Using these criteria, the experimenter's wife made independent judgments of recognition versus nonrecognition for the total of 2,100 responses from the two groups of subjects. There was less than 3% disagreement with my scoring, with the disagreements showing no consistent directional bias in favor of either the schizophrenics or the normals.

[9] Since only males were used as subjects in this experiment, the conclusions based on these data cannot be generalized to women. In this connection, it is of interest to note that Rosenstock (1951) has presented some experimental evidence indicating that women tend to repress aggressive material, while men tend to repress sexual material.

Each subject was shown a series of 24 pictures representing various aspects of sexuality (sexual intercourse, birth, the genitalia, etc.) and he was asked to describe what he saw at .01-second presentations of the stimuli. As a control against the possibility that the responses given to these stimuli were not a specific function of the nature of the ideational areas represented, 9 blank cards and 18 nonsexual pictures were also presented. The prerecognition distortions given to samples of both sexual and nonsexual stimuli were rated by independent judges as either "sustaining" or "nonsustaining" images, with the judges blind as to whether a response was given by a schizophrenic or a normal individual.

The quantitative experimental results support the following conclusions:

1. While both normal males and schizophrenic males in our culture are threatened by some aspects of sexuality, schizophrenics are threatened by more aspects of sexuality than are normals, and this differential in threat is not simply a function of the schizophrenics' indiscriminate fear of all kinds of objects and activities.

2. The indirect representation of sexual stimuli often occurs when these stimuli are tachistoscopically presented just below the time threshold for conscious recognition and report.

3. There are discriminable differences between normal males and schizophrenic males in the indirect representation of sexual stimuli under the conditions of this experiment. The schizophrenic group significantly exceeded the normal group in the production of nonsustaining imagery (reflecting negative affective attitudes) in response to sexual stimuli, while normal subjects significantly exceeded schizophrenics in the production of sustaining responses (reflecting positive affective attitudes) to these stimuli.

## REFERENCES


AREV, L. B. A study in perception and symbolism. Unpublished doctoral dissertation, University of Chicago, 1952.

BLEULER, E. *Dementia praecox or the group of schizophrenias*. New York: International Univer. Press, 1950.

CAMERON, N. The functional psychoses. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol. 2. New York: Ronald, 1944. Pp. 861–921.

FISHER, C. Dreams and perception. *J. Amer. Psychoanal. Ass.*, 1954, **2**, 389–445.

FISHER, C. Dreams, images and perception: A study of unconscious-preconscious relationships. *J. Amer. Psychoanal. Ass.*, 1956, **4**, 5–47.

JONES, E. The theory of symbolism. In E. Jones, *Papers on psychoanalysis*. (5th ed.) Baltimore: Williams & Wilkins, 1949. Pp. 87–144.

POETZL, O. Experimentelle erregte Traumbilder in ihren Beziehungen zum indirekten Sehen. *Zbl. ges. Neurol. Psychiat.*, 1917, **37**, 278–349.

ROSENSTOCK, I. M. Perceptual aspects of repression. *J. abnorm. soc. Psychol.*, 1951, **46**, 304–315.

SHEVRIN, H., & LUBORSKY, L. The measurement of preconscious perception in dreams and images: An investigation of the Poetzl phenomenon. *J. abnorm. soc. Psychol.*, 1958, **56**, 285–294.


(Received September 14, 1959)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 432–439

# THE UNCONSCIOUS PERCEPTION OF THE MEANING OF VERBAL STIMULI[1]

MARCUS J. FUHRER AND CHARLES W. ERIKSEN
*University of Illinois*

WITHIN the last 10 years, a number of experiments (Blum, 1955; Dixon, 1956, 1958; Lazarus & McCleary, 1951) have attempted to demonstrate the operation of a supersensitive perceptual system functioning below the individual's awareness threshold. At the speculative level, Osgood (1957) has suggested that "certain dominant components of the mediating reaction to a sign can be elicited at stimulus levels below that at which other components occur" (p. 396). "Dominant evaluative (emotional) components" are included by Osgood in that class of mediating responses which may occur temporally prior to accurate recognition of the stimulus (p. 397). Lazarus (1956) has indicated that the dimension of threat versus nonthreat may be one of these subliminally discriminated, evaluative components.

Eriksen (1956a, 1956b, 1958) has formulated an alternative theoretical model which assumes neither prior nor supersensitive discrimination, but which accounts for correct discriminative responses in the absence of verbalized discrimination. The theory conceptualizes perception as a mediating process, occurring between stimulus input and response output. The perceptual process is assumed to give rise to a number of different responses, of which verbal identification is only one. As applied to subception-like effects, the model presupposes concurrent responses to a stimulus, e.g., a verbal response and an affective response, both of which are less than perfectly correlated with the stimulus and which possess independent sources of error. In response to impoverished stimulation, cases thus may arise in which the affective response is correct when the identification is inaccurate and vice versa. In contradistinction to the theories of Osgood and Lazarus, this formulation neither assumes that the affective responses occur temporally prior to verbal identification nor that the threshold for affective discrimination is lower than that for verbal identification. The only necessary conditions for obtainment of a subception-like effect is that both the verbal identification and the affective responses are responding to the mediating perceptual process and have at least partially independent sources of error.

The concurrent response model and the theorizing of Osgood and Lazarus have recently been put to a differential test by Eriksen, Azuma, and Hicks (1959). In this experiment, two sets of words, previously judged to be pleasant or unpleasant, were presented tachistoscopically to a group of subjects (*S*s). Following exposure of each word, the *S*s judged whether the word was pleasant or unpleasant and also guessed what the word was. The judgments of affective classification that occurred in the absence of correct identification of the word then were examined for above chance accuracy. The authors report that in the absence of correct identification of the stimuli, the *S*s' affective judgments were correct significantly more often than could be attributed to chance. However, as opposed to the theories of Osgood and Lazarus, no evidence was found that the affective judgments had a lower discrimination threshold than the identification responses.

An experiment by Dixon (1958) is of particular interest because it generated evidence of unconscious discrimination which appears relatively immune from alternative interpretation by the concurrent response model. Dixon hypothesized that subliminal stimuli would activate autistic verbal behavior which could be related meaningfully to the subliminal stimuli. Dixon presented a series of predominantly verbal stimuli at an illumination level so low that the *S*s reportedly were unable to discriminate even the presence of the stimuli. The *S*s' task was to guess the identity of the stimulus items. That the guesses were actually associations to the stimuli was shown

[1] This research was supported by Research Grant M-1206 from the National Institute of Mental Health, United States Public Health Service.

by the *Ss*' better than chance ability subsequently to indicate the stimulus item to which each guess had been given. Because of these surprising results, and because Dixon explicitly labels his experiment a "pilot study," the present investigation was initiated to explore the possibilities inherent in Dixon's interesting method and possibly to extend his findings.

## METHOD

### *Design*

Four treatment groups were employed in the present investigation. In all groups, *Ss* were run individually and participated in two separate sessions. The first session was divided into two stages: threshold determination and subsequent presentation of the stimuli. In the second session, held 3 days later, each *S* was presented with his responses made during the first session and asked to match these against the original stimulus items. Procedural variations among the four groups differed primarily in terms of how the thresholds were obtained, the level of illumination at which the stimulus items were presented, and the number of items to which the *S* matched his original responses.

The procedure used in Group I was directly comparable to that used by Dixon, except that GSRs were not recorded. The instructions also were changed slightly to increase the probability that the *Ss* would not hesitate to report if anything was perceived during the presentation of the stimuli. While Dixon employed only 7 *Ss*, 16 *Ss* were used in our Group I.

To obtain threshold determinations for *Ss* in Group II, a forced-choice technique was used. This technique has been shown by Blackwell (1953) to yield more reliable threshold values than techniques of the kind used by Dixon. Whereas the stimuli for Group I were presented well below the *S*'s previously determined threshold, the same stimuli were presented for Group II slightly above the obtained thresholds. It was anticipated that comparison of the results from Groups I and II might reveal a range of illumination values at which autistic responses to stimuli could be obtained.

Group III was employed to determine how much the *Ss*' responses were influenced by structural characteristics of the stimuli, e.g., the height of the letters and the length of the words, as opposed to the meaning of the stimuli. For *Ss* in Group III, a threshold was obtained such that the *S* could accurately identify approximately 15% of a series of test words. To determine the relative influence of structural cues, the *Ss* matched their guesses to a list containing the actually presented stimulus items, as well as 10 other words of similar structural composition, but of reversed emotional connotation. For example, if the actual stimulus was "male sex organ," the *S*'s list included this phrase, as well as the affectively more neutral phrase "made six orders." Emotionally neutral stimuli also had structurally similar, but emotionally connoted counterparts.

A fourth treatment group was employed primarily to determine how response biases affect the *Ss*' matchings under the conditions used in Group III. The stimuli were presented upside down and backwards, thus ruling out the possibility of an *S* perceiving the meaning of the verbal stimuli. By this method of presentation, structural cues were confounded, with the exception of word or phrase length.

### *Subjects*

The *Ss* were undergraduate students from the introductory psychology course at the University of Illinois. They were required to serve a minimum number of hours as experimental *Ss* as part of their course requirements. Each of the 64 *Ss* was randomly assigned to one of four treatment groups, consisting of 8 males and 8 females.

### *Apparatus*

To minimize uncontrolled cues from the experimenter (*E*), the *S* was seated in a wooden booth 6 ft. by 6 ft. by 4 ft. The only openings were a door and an aperture 11 in. by 7 in. in the far wall 38 in. from the floor. The aperture, through which the stimuli were projected, was covered with a ground glass screen to give a homogeneous field illuminated with a brightness of .3 log ft-L. A chair was provided, as well as an adjustable chin-rest 38 in. from the floor and positioned 44 in. from and directly in front of the screen.

The stimulus items were photographed on 35-mm. film and placed on individual slides, masked except for the exposed portion of the film. An automatic projector, supplied by a constant voltage transformer through a variac, was used to project the stimuli. Image brightness was controlled by a specially calibrated projector diaphragm, making increments of .05 log ft-L possible throughout the range of illumination used in the experiment. A shutter, manually operated by *E*, was used to expose the stimuli. Opening the shutter automatically activated a chronotron; closing the shutter automatically stopped the chronotron.

A standard set of stimuli were used in the four treatment groups. There were four "emotionally neutral" words and six words of "primary emotional interest." The items were presented in the following order: this, penis, barn, father, vagina, seven, male sex organ, line, mother, female sex organ. For Group I, this standard set of items was increased to include a horizontal and a vertical line, positioned second and fifth, respectively, in the list. For Group II, the stimulus item house was used to make threshold determinations. For Groups III and IV, threshold determinations were made using the items: country, time, other, almost, present, because.

### *Procedure*

The procedure described for Group I was the same in each of the other three groups, with the exception of specific variations described below in detail. All *Ss* were run individually in two separate sessions. The first session was divided into two stages, threshold determination and subsequent presentation of the stimuli.

*Group I.* The *S* was seated in the darkened booth and instructed to relax during an 8-min. period of dark adaptation. A clearly visible horizontal line then was projected onto the center of the screen, so as to subtend a visual angle of 2° 48′. The *S* was asked to position his chin on a chin-rest, and these instructions were read:

> You see before you, near the center of the screen, a horizontal line of light. I am going to make the line gradually dimmer. All you have to do is to watch it carefully and say "Stop" the moment it has completely disappeared. Don't say "Stop" until you are quite certain you cannot see anything at all upon the screen.

The brightness of the line was diminished at the rate of .1 log ft-L/sec., until the *S* said "Stop." A series of judgments was obtained also using a vertical line. The *S*'s threshold was defined as the lowest illumination level which was not equaled or bettered in three subsequent trials, considering both series of judgments.

> Introductorily to the presentation of the stimuli, the *S* was instructed: I am going to project words on the screen but often they may be too dim for you to see. With each projection, however, I want you to make a guess regarding identity of the word shown. Your signal that a word is there and that you should respond will be my saying "Now." During the experiment, keep watching the screen even though you may see little upon it. Try and maintain a relaxed, passive attitude, just allowing each guess to be the first word that comes to mind after I say "Now." I will give you no indication of whether you were right or wrong.

Ten practice responses were obtained, without a stimulus word being presented. The projector diaphram then was adjusted to give a level of illumination .3 log ft-L below the *S*'s previously determined threshold. Simultaneously with the opening of the shutter, *E* signaled "Now," the timer was automatically started, and the first stimulus word was presented. Immediately following *S*'s response, the shutter was closed, and the timer simultaneously stopped. The *S*'s response latency and verbal response were recorded by *E*. The *S* was asked if he had seen anything on the screen. This procedure was followed for the presentation of each of the 12 stimuli. The stimuli were presented in the same order three more times, without indication that the series was being repeated, until four responses were obtained to each of the 12 stimuli. The *S* then was removed from the booth and was asked to associate to each of the responses he had given.

A second session was held 3 days later. A card was handed the *S* upon which were typed in random order the stimuli used in the first session. The *S* was asked if the list seemed familiar in any way. The following instructions then were read:

> I am going to read to you a list of words. After each I want you to tell me the number of that item on your list which in your opinion is most appropriate to the word I have given. I shall not ask you for the reasons for your choice which should be entirely personal and not bound by logical or common-sense considerations; just choose the item which you feel, for any reason whatever, is most suited to the word I have given. Take as much time as you like, then give me the number of your choice.

The *S*'s original responses were read in random order, and the *S*'s matches were recorded.

*Group II.* The *S*'s threshold was determined by a forced-choice technique. The following instructions were given:

> I am going to present a series of two presentations on the screen. In one of the two presentations, I will present nothing, while in the other, I will present a dimly illuminated word. Your task is to guess whether the word appeared in the first or the second presentation. I will say "the first" to signal the first presentation, pause 4 sec., and then say "the second" to signal the second presentation. I then will ask you for your guess. I will give you no indication of whether you were right or wrong. We will proceed then to the next set of presentations in the series.

A series of presentations was given; the position of the word *house* was randomly ordered within each set of presentations. The level of illumination was increased at increments of .2 log ft-L until *S* made seven out of eight correct judgments. This illumination level then was used to present 10 of the 12 stimuli used in Group I, the vertical and horizontal lines being excluded. In all other respects, the procedure did not vary from that used in Group I.

*Group III.* The *S*s' thresholds were obtained using the following instructions:

> I am going to present a series of different words on the screen. Your task is to guess the identity of the word shown. When I give the signal "Now," a word will be presented, and you are to announce your guess aloud. In some cases, the word may appear so dimly that you feel your guess is rather arbitrary. Don't let that disturb you, for you may be guessing more accurately than you think. I will give you no indication whether your guess was right or wrong.

The series of test words was repeated in random order at increments of .2 log ft-L until the *S* could accurately identify one of the seven words. At this illumination level, the standard set of stimuli were presented, with the prior instructions given:

> Your task remains the same, i.e., to guess the identity of the words presented on the screen. Don't be disturbed if you feel that some of your guesses are rather arbitrary. Just try to make as accurate a guess as possible with the information you get from the screen. I will give you no indication whether your guess was right or wrong.

The subsequent procedure used in the first session did not vary from that used for Groups I and II, with the exception that the *S* was not asked whether or not he had seen the stimulus.

In the second session, the *S*s matched their original responses to a list of items containing not only the 10 original stimuli, but also 10 additional items. Each of the additional items was a structural counterpart of one of the original stimuli, but was reversed in affective connotation. The original stimuli, with their counterparts indicated in parentheses, were as follows: barn (harm), seven (sexed), this (piss), line (lied), male sex

organ (made six orders), vagina (vague), mother (matter), female sex organ (former sea origin), father (fatter), penis (pens).

*Group IV.* The thresholds for these *S*s were obtained in the same way as for Group III. The subsequent procedure differed in that the 10 stimuli were presented upside down and backwards on the screen, .3 log ft-L below the *S*'s previously determined thresholds. At this illumination level, the *S*s could clearly perceive the presence of the stimuli, but they did not detect the unusual orientation of the stimuli.

## RESULTS AND DISCUSSION

### *Groups I and II*

Because of similarity in experimental procedure, the results from Dixon's study and from Groups I and II may be compared directly. The thresholds for *S*s in Group I were determined by the procedure previously used by Dixon. Because of the possible unreliability of the thresholds obtained by Dixon's method, thresholds for *S*s in Group II were determined by a forced-choice method, and the stimuli were presented at a slightly higher illumination level than employed in Group I.

To determine if the *S*s could match their responses with the stimulus to which each response had been made originally, we followed Dixon in using a statistical test developed by Stevens (1939).

Four of the 16 *S*s in Group I made more than the chance number of correct matches, though none of the 4 was significant at the .05 level. One *S* obtained no correct matches, a result which is significant beyond the .02 level, in a direction opposite from that which was predicted. Dixon reports that 5 of his 7 *S*s made more correct matches than chance expectation, with 2 achieving significance beyond the .05 level. Two of Dixon's *S*s scored below chance, but neither score was statistically reliable.

The Stevens test permits the deviations from chance and the variances to be added over a number of *S*s to provide a significance test for the group's performance. Using this analysis for the data from Group I, a $z$ value of $-1.062$ was obtained, which, for a one-tailed test, was not only insignificant ($p > .14$), but in the opposite direction from Dixon's results. Application of a correction for discontinuity, suggested by Mosteller and Bush (1954), resulted in a $z$ value of $-1.128$ ($p > .13$). Applying the same analysis, without the correction for discontinuity, Dixon obtained a significant $z$ value of 2.57 ($p < .005$).

While our Group I *S*s were unable to match meanings to the specific stimuli at better than chance level, the possibility existed that they might be able to make dichotomous discriminations between emotionally charged and neutral stimuli. To make this determination, a fourfold chi square was calculated for each *S* to determine the relationship between the affective classification of the stimulus matches (emotional or neutral) and the affective classification of the original stimulus. None of the 16 chi squares reached significance at the .05 level. Since three of the *S*s obtained values of chi square that reflected a relationship in the opposite direction, the over-all significance of the results could not be established by the usual method of adding independent chi square values. Instead, the Fisher-Pearson test was used (Gordon, Loveland, & Cureton, 1952). This test involves computing the exact probability for each obtained chi square value and then transforming them to $\log_e$. The sum of these transformed values is distributed as chi square with $2n$ *df* or, in the present case, 32 *df*. The application of this test to the data yielded a summed chi square value of 31.78, which clearly failed to attain significance ($p > .40$).

Following Dixon's analysis, response latencies were examined to determine if shorter reaction times were given to emotionally neutral stimuli. Of the 16 *S*s in Group I, 7 manifested shorter average latencies for the neutral stimuli. In comparison, 5 of Dixon's 7 *S*s had shorter average latencies for the neutral items. Neither of these results was significant using the sign test. For Group I, a nonsignificant $t$ value of .7902 (15 *df*, $p > .20$) was obtained for the difference of the average response latencies to neutral versus emotional items.

To determine if sex differences were present, all the above analyses were repeated for the eight male and eight female *S*s, but no significant sex differences were revealed.

Next, Group II: Eight *S*s made more than the chance number of correct matches, two of which approached or exceeded significance ($p < .059$ and .015). On the other hand, three of the scores falling below chance had $p$ values

of .054, .048, and .017. Application of the Stevens test to the summed data from Group II resulted in a nonsignificant $z$ value of .3354, corrected for continuity ($p > .35$).

Of the 16 chi square values relating the affective classification of the stimuli and the *S*s' stimulus matches, 8 were in the predicted direction, though none was significant at the .05 level. The Fisher-Pearson test yielded a chi square value of .2698 which, with 32 *df*, was not significant ($p > .50$). No significant differences were found for the reaction times of Group II to emotional or neutral stimuli, nor were there any significant differences attributable to sex.

The results from Groups I and II were consistent in failing to provide evidence that *S*s can perceive meaning at exposure levels too low for the *S*s to report the presence of the stimuli or at levels at which the *S*s can report the presence of the stimuli but cannot identify the stimuli.

*Groups III and IV*

While we were unable to confirm Dixon's findings that meanings can be perceived well below the recognition threshold, the possibility remained that meanings could be perceived prior to recognition, but where the gap between the presence-absence threshold and the recognition threshold was considerably less than used by Dixon. To test this possibility, Group III was employed.

Since the stimuli for this group were exposed at an illumination level at which 13% of the stimuli were correctly identified at the time of exposure, only those responses were analyzed in which the stimuli were incorrectly identified. In addition, 37 of the 128 responses to the stimulus phrases male sex organ and female sex organ were excluded because the *S* had correctly identified one or more of the words in these phrases.[3]

Application of the Stevens test to the reduced data from Group III revealed that 12 of the 16 *S*s made more correct matches than would be expected by chance.

Two of the *S*s' scores were significant beyond the .05 level, with one other significant beyond the .0001 level. None of the *S*s' scores which fell below chance expectation were statistically significant. Application of the Stevens test to the summed data from Group III yielded a $z$ value of 3.61, corrected for continuity, which is significant beyond the .0002 level. This analysis revealed that when illumination levels were substantially higher than used in Groups I or II or by Dixon, *S*s were able to correctly match their initially incorrect responses with the stimuli which produced these responses. But, before attributing the better than chance performance of Group III to the unconscious perception of meaning, the results from Group IV must be examined.

It will be recalled that *S*s in Group IV responded to stimuli presented upside down and backwards at an illumination level only slightly lower than employed in Group III. Eight *S*s in Group IV obtained a number of correct matches which exceeded the chance expectation. Three of these scores were significant beyond the .05 level, while none of the scores falling below chance expectation were significant. The combined scores for Group IV, corrected for discontinuity, yielded a $z$ value of 2.18, significant beyond the .02 level. The analysis revealed that even when the dimly illuminated stimuli were presented upside down and backwards, *S*s were able to match their incorrect identifications to a better than chance extent. This suggested that structural cues, e.g., differences in length among the stimulus words and phrases, must have been responsible for Group IV's performance, since the method of presentation, in all likelihood, precluded the perception of the meaning of the stimuli.

To determine if Group III *S*s were responding to the meaning of the stimuli in addition to structural cues, it was necessary to determine if Group III performed significantly better than Group IV. A nonsignificant $t$ value of .51 was obtained (30 *df*, $p > .30$). Since the two groups did not differ significantly in their number of matches, it would appear that the performance of Group III is adequately explained in terms of nonmeaning cues such as

[3] If the *S* had perceived the word female in the phrase female sex organ, he had a cue in addition to meaning with which to improve the probability of his subsequently making a correct match. That initially correct identifications were highly determinate of the *S*s' subsequent matches is indicated by the finding that 95% of the correct identifications and 75% of the partial identifications led to correct matches.

word or phrase length. To check this conclusion and to rule out the implausible possibility that Group IV *Ss* actually were perceiving the meaning of upside down and backward stimuli, another analysis was employed.

It will be recalled that for Groups III and IV, the list of items to which the *S* was to match his responses contained not only the 10 stimuli actually presented, but also 10 additional items. Each of the additional items was a structurally similar counterpart of one of the original stimuli, but it was reversed in emotional connotation. If the *S* had originally responded to the meaning of the stimulus, his match could be expected to be to the actual stimulus to which the response had been made. On the other hand, if the *S* had responded to the structural characteristics of the stimulus, he would be expected to show an equal preference in his matches either for the correct stimulus or for its structurally similar counterpart. Therefore, for each *S*, we computed not only the number of matches to the correct stimulus, but also the number of matches to its structural counterpart. A difference score was obtained for each *S* by subtracting the number of matches made to structurally similar items from the number of matches made to the correct stimuli. The mean difference score of .125 obtained for Group III was positive in sign, but did not differ significantly from 0 ($t = .307$, 14 *df*, $p > .7$).

The above result offered no evidence that the *Ss* were perceiving meaning in the absence of correct identification of the stimulus. There was a possibility, however, that the *Ss*' response biases may have masked any effect attributable to the unconscious perception of meaning. The *Ss* may have been reluctant to verbalize the emotional phrases and words as matches to their responses and, instead, chose the emotionally more neutral but structurally similar phrases. To evaluate the extent to which response bias may have operated in the data, we again compared the performance of Groups III and IV. For each *S* in each group, the difference score, i.e., the number of correct stimulus matches minus the number of matches to the structurally similar counterparts, were computed separately for the four neutral and six emotional stimuli. These scores then were analyzed by a modified three-way classification analysis of variance (groups, emotional versus neutral stimuli, and individuals).

TABLE 1

MEAN DIFFERENCE SCORES BY GROUPS AND AFFECTIVE CLASSIFICATION

| | Emotional Stimuli | Neutral Stimuli |
|---|---|---|
| Group III | −1 | 1.125 |
| Group IV | −1.063 | .375 |

Only the main effect attributable to the affective classification of the stimuli was significant in the analysis ($F = 15.91$, 1 and 30 *df*, $p < .01$). The nature of these results can be seen in Table 1 which shows the mean of the difference scores by groups and affective classification. In both groups, the *Ss* tended to choose the emotionally neutral member of the pair in making their matches. Thus, when an emotional stimulus had been responded to, the *Ss* tended to match their response to the structurally similar, but emotionally neutral control. In the case in which a neutral stimulus had been present, the *Ss* again tended to match their response to the neutral alternative, which in this case was the correct stimulus. It is clear from this analysis that the differential stimulus-matching of responses to emotional and neutral stimuli was attributable to the *Ss*' bias to choose affectively neutral items for their matches.

The possibility may be raised that response biases masked the *Ss*' ability to discriminate emotional items at the level of exposure used. This alternative is contradicted by the finding of no significant interaction of Groups III and IV with emotional or neutral words. As can be seen, the *Ss*' response bias, while a significant effect, is not overriding (a mean difference of −1 for emotional stimuli, versus a possible difference of −6 if response bias was the sole determiner of choice). Thus, if perception of meaning was occurring, it should have mitigated the bias such that Group III would more accurately match responses made to emotional stimuli than would Group IV. However, as can be seen, the means are virtually identical offering no suggestion that perception of meaning was influencing these matches.

In addition, this analysis serves to render highly improbable the suggestion that Group IV *Ss* were also perceiving meaning. Even if

Group IV *S*s were perceiving the meaning of the upside down and backward stimuli, it is to be expected that their perception of meaning should be less accurate than for Group III *S*s. Yet, Table 1 reveals this was not the case.

Our results are consistent in providing no evidence for the hypothesis that *S*s perceive the meaning of verbal stimuli at durations or illumination levels too low for verbal identification. Therefore, the results are quite at variance with Dixon's report. The question arises as to why the discrepancy between his results and the present findings.

First, the problem of sampling fluctuations must be considered. In Dixon's study, only seven *S*s were employed, of whom only five gave evidence of the unconscious perception of meaning. A second likely basis for the discrepancy between the two studies lies in the method of threshold determination employed by Dixon and the instructions given his *S*s. He used a modified method of limits and instructed his *S*s: "I am going to project words on the screen but they will be too dim for you to see" (Dixon, 1958, p. 30).

As has been amply demonstrated, threshold measurements depend importantly upon the instructions given the *S*s as well as the method employed to obtain the thresholds (Blackwell, 1953; Eriksen, 1958; Goldiamond, 1958). Dixon's instructions may have had the effect of giving the *S*s a set for not reporting the possibly perceived presence of the stimuli. The possibility exists that while Dixon's *S*s were unable to recognize the various words or even to perceive the letters, clearly, they nonetheless were able to pick up such structural cues as differences in word length or phrase length. These length cues were particularly marked in the set of stimuli employed in both Dixon's and the present study. That *S*s can utilize such cues is clearly demonstrated by the *S*s in our Group IV. For this group the verbal stimuli were presented upside down and backwards to eliminate all cues except differences in stimulus length. Yet, this group showed significantly better than chance matching.

Dixon fails to report whether the *S*s were sufficiently dark adapted before thresholds were obtained. It is possible that as dark adaptation progressed, Dixon's *S*s attained a degree of perceptual sensitivity not reflected in the thresholds obtained at the first of the experimental session. This, together with his rather inexact method of threshold measurement may account for the discrepancies between his results and ours.

## SUMMARY

The hypothesis was investigated that the meaning of a verbal stimulus may be responded to without prior recognition of the stimulus. Four treatment groups were employed of 16 *S*s each. The experimental procedure used in the first two groups was very similar to that previously used by Dixon. The procedure consisted essentially of obtaining the *S*'s guesses to a series of verbal stimuli, dichotomized in affective connotation, which were presented at or slightly below the *S*'s previously determined absolute threshold. The results from these two groups were consistently at variance with results obtained by Dixon, for there was no evidence that the *S*s could make better than chance matches.

In the third treatment group, the same stimuli were presented at a higher level of illumination, such that 13% of the *S*s' guesses were correct identifications of the stimuli. The list of items to which the *S*s subsequently matched their guesses contained not only the 10 original stimuli, but also 10 additional items that were structural counterparts of the original stimuli, but reversed in affective connotation. A fourth treatment group made guesses to stimuli that were presented upside down and backwards to preclude the possibility that responses were made to the meaning of the stimuli. While both groups were able to make better than chance matches, the groups' performances did not significantly differ. Both groups were predominantly responding to structural differences among the stimuli, particularly to the markedly different length of the stimuli. It was concluded that no evidence was obtained that *S*s respond to the meaning of verbal stimuli prior to recognition of the stimuli.

## REFERENCES

BLACKWELL, H. R. Psychophysical thresholds: Experimental studies of methods of measurement. *Bull. Engng. Res. Inst. U. Mich.*, 1953, **36**.

BLUM, G. S. Perceptual defense revisited. *J. abnorm. soc. Psychol.*, 1955, **51**, 24–29.

DIXON, N. F. Symbolic associations following subliminal stimulation. *Int. J. Psycho-Anal.*, 1956, **37**, 159–170.

DIXON, N. F. The effect of subliminal stimulation upon autonomic and verbal behavior. *J. abnorm. soc. Psychol.*, 1958, **57**, 29–36.

ERIKSEN, C. W. An experimental analysis of subception. *Amer. J. Psychol.*, 1956, **69**, 625–634. (a)

ERIKSEN, C. W. Subception: Fact or artifact? *Psychol. Rev.*, 1956, **63**, 74–80. (b)

ERIKSEN, C. W. Unconscious processes. In M. R. Jones (Ed.), *Nebraska symposium on motivation*. Lincoln: Univer. Nebraska Press, 1958.

ERIKSEN, C. W., AZUMA, H., & HICKS, R. B. Verbal discrimination of pleasant and unpleasant stimuli prior to specific identification. *J. abnorm. soc. Psychol.*, 1959, **59**, 114–119.

GOLDIAMOND, I. Indicators of perception: I. Subliminal perception, subception, unconscious perception: An analysis in terms of psychophysical indicator methodology. *Psychol. Bull.*, 1958, **55**, 373–405.

GORDON, M. H., LOVELAND, E. H., & CURETON, E. E. An extended table of chi-square for two degrees of freedom, for use in combining probabilities from independent samples. *Psychometrika*, 1952, **17**, 311–316.

LAZARUS, R. S. Subception: Fact or artifact? A reply to Eriksen. *Psychol. Rev.*, 1956, **63**, 343–347.

LAZARUS, R. S., & MCCLEARY, R. A. Autonomic discrimination without awareness: A study of subception. *Psychol. Rev.*, 1951, **58**, 113–122.

MOSTELLER, R., & BUSH, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *The handbook of social psychology*. Vol. I. Cambridge, Mass.: Addison-Wesley, 1954.

OSGOOD, C. E. Motivational dynamics of language behavior. In M. R. Jones (Ed.), *Nebraska symposium on motivation*. Lincoln: Univer. Nebraska Press, 1957.

STEVENS, W. L. Tests of significance for extrasensory perception data. *Psychol. Rev.*, 1939, **46**, 142–150.

(Received September 21, 1959)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 440–447

# THE CHILD'S DIFFERENTIAL PERCEPTION OF PARENTAL ATTRIBUTES[1]

JEROME KAGAN AND JUDITH LEMKIN

*Fels Research Institute, Yellow Springs, Ohio*

THE child's conceptualization of the parental roles and the specific characteristics he attributes to each parent are relevant to several theoretical issues. In describing the course of the child's identifications during the early school years, it has been suggested that the choice of parental models for identification is determined by the child's perception of parental differences in power, competence, and nurturance (Kagan, 1958; Maccoby, 1959). Moreover, Parsons and Bales (1955) have suggested that the mother's role is primarily concerned with maintaining warm, integrated, interpersonal relations. The paternal role prescription calls for instrumental skills and the ability to adapt to the environment. Thus, information on the child's perception of his parents may clarify problems associated with the identification process and test the validity of Parsons' dichotomy of parental roles.

Semantics is a second area which draws its theoretical statements from the individual's perception of the environment. Osgood, Suci, and Tannenbaum (1957) have reported that young adults use three conceptual dimensions in their categorization of a wide variety of objects, i.e., evaluative (good vs. bad), potency (strong vs. weak), and activity (active vs. passive). The content of the potency and activity dimensions resembles the stereotyped conception of male and female sex roles. For example, the adjectives large, strong, loud, heavy, brave, bass, rough, and rugged have the highest factor loadings on the potency factor. The adjectives agitated, sharp, ferocious, tense, hot, angular, active, and fast have the highest loadings on the activity factor. These adjectives are more commonly applied to males than to females, while their polar opposites are more characteristic of stereotyped female attributes.

[1] This research was supported, in part, by research grant (M-1260) from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service.

Since the three factors discovered by Osgood et al. are relatively invariant over a variety of populations, it seems profitable to ask about the developmental history of this language learning. It is possible that these ingrained modes of conceptualization are a function, in part, of the child's perception of parental attributes. If a typical child perceives the father as stronger and more active than the mother, then one might expect the adult's conceptualization of "maleness" and "femaleness" to be colored by these earlier perceptions. Jenkins and Russell (1958) asked 540 college students to use the semantic differential in rating a variety of concepts, including mother and father, on 20 polar opposite adjectives. These young adults rated father, in relation to mother, higher on the dimensions cruel, successful, strong, and bad. It would be of interest to determine how early this pattern of labels is learned and whether the 6-year-old characterizes his father in a similar way.

This paper summarizes a study on the perception of parents by children 3 to 8 years of age. Several years ago, the senior author interviewed children 6 to 10 years of age and found that the father was perceived as more dominant and less friendly than the mother (Kagan, 1956). Gardner (1947) found similar results for children 10 to 13 years of age. Emmerich (1959) used a more indirect procedure with 4- and 5-year-olds and reported that children viewed the mother as more nurturant and less dominating and interfering than fathers.

In the Kagan study the questions were so phrased that the child was answering for himself and his parent, e.g., "Who punishes you, your mother or your father?" In the Emmerich study the questions were indirect and the child was responding in terms of "make believe" parents and children. No one has systematically studied the relationship between these two procedures. The present research had two purposes: to study the differential perception of mothers and fathers by

young children with respect to power, punitiveness, competence, and nurturance; and to compare the protocols obtained by both direct and indirect methods of questioning.

## METHOD

### *Subjects*

The subjects were 32 boys and 35 girls ranging in age from 3–9 to 8–6 with a median age of 5–6 for the group. The children were enrolled in one of three nursery schools or the first grade of an elementary school at the time of the testing. Each child was tested individually at the school by the junior author (the experimenter). Several days prior to testing the experimenter visited the school in order to become acquainted with the subjects and reduce the effect of stranger anxiety on the test performance.

### *Procedure*

Each child was asked four sets of questions in a standard order. The first set of questions involved color vs. form preferences in a simple concept formation task and is not relevant for the present report. Following this task the child was asked the remaining three sets of questions.

*Set A. Perception of parents—indirect method.* Initially, the experimenter showed the child a drawing of a family scene containing a father, mother, boy, and girl and said: "Here is a picture of a mommy, daddy, a boy, and a girl." She then replaced this family picture with two separate illustrations of a mother and a father and said: "Here are two pictures of just the mommy and daddy. Which one is the mommy? Which one is the daddy? I am going to ask you some questions about the mommy and daddy. You point to the one you think." The questions follow:

1. Who is the strongest one?
2. Who gives (the boy) (the girl) the most presents?
3. Who spanks the most?
4. Who is the boss at home?
5. Who does the (boy) (girl) like the best?
6. Who gets mad at the (boy) (girl)?
7. Who is the smartest?
8. If the (boy) (girl) did something bad and both mommy and daddy were home, who would punish the (boy) (girl)?
9. Who is nicer to the (boy) (girl)?
10. Who is the (boy) (girl) more scared of?
11. Who tells the (boy) (girl) what to do all the time?
12. Who kisses the (boy) (girl) the most?
13. Who would the (boy) (girl) like to grow up to be, who would the (boy) (girl) want to be like when (he) (she) grows up?

*Set B. Perception of parents—indirect picture method.* The experimenter then showed the subject a series of pictures (separate sets for boys and girls) in which a boy or a girl was in a situation that normally involved a parent figure but in which no adult was illustrated.[2] In some cases the hands of an adult were illustrated but no identification related to the sex of the adult was evident. The experimenter then said: "Here are some pictures which show a (boy) (girl) but the mommy and daddy are missing. You tell me who is missing—the mommy or the daddy." Descriptions of the pictures and the experimenter's explanatory statement to the child follow:

1. Some hands are shown giving the child an ice cream cone.

*E:* "Someone is giving the (boy) (girl) an ice cream cone—who is it? The mommy or the daddy?"

2. Some hands are shown reaching up for a toy that the child cannot reach.

*E:* "Someone is getting the (boy) (girl) the toy (he) (she) can't reach. Who is getting it?"

3. Some hands are shown spanking the child.

*E:* "Someone is spanking the (boy) (girl). Who is spanking? The mommy or the daddy?"

4. The child is held up high in the air by a pair of hands and has a smile on his face.

*E:* "Someone is picking the (boy) (girl) up in the air. Who is picking (him) (her) up? The mommy or the daddy?"

5. The child is cowering.

*E:* "Someone is yelling at the (boy) (girl). Who is yelling at the (boy) (girl)?"

6. Some hands are illustrated handing a gift wrapped box to the child.

*E:* "Someone is giving the (boy) (girl) a present. Who is giving (him) (her) a present?"

7. A dog is chasing a child who looks frightened and is running from the dog.

*E:* "The dog is chasing the (boy) (girl) and (he) (she) is running to someone for help. Who is the (boy) (girl) running to?"

8. The child is holding the hand of someone.

*E:* "Someone is taking a walk with the (boy) (girl). Who is it? The mommy or the daddy?"

9. The child is on a seesaw.

*E:* "Someone is on the seesaw with the (boy) (girl). Who is it? The mommy or the daddy?"

10. The child is sticking out his tongue with an angry expression.

*E:* "The (boy) (girl) is mad at someone. Who is (he) (she) mad at? The mommy or the daddy?"

*Set C. Perception of parents—direct questioning method.* After the pictures were administered the experimenter repeated the 13 questions that were asked in Set A, but the questions were asked in a different order. This time the experimenter phrased the question so that the child was talking about himself and his own parents. The experimenter said: "Now I am going to ask you a few questions about your mommy and your daddy." For example, the first question was phrased: "Who gives you the most presents? Your mommy or your daddy?"

The child's answers to all questions were recorded on a prepared form. If the child replied with neither mother nor father, an inquiry was made by the experimenter. If the child persisted in giving no answer or an answer other than mother or father the experimenter proceeded on to the next question. The majority of children understood the task and responded with either mother or father to all questions.

[2] Photostats of these pictures are available upon request.

TABLE 1

PERCENTAGE OF BOYS AND GIRLS ANSWERING MOTHER OR FATHER TO INDIRECT AND DIRECT QUESTIONS

| Question | Indirect | | | | Direct | | | |
|---|---|---|---|---|---|---|---|---|
| | Mother | | Father | | Mother | | Father | |
| | Boys | Girls | Boys | Girls | Boys | Girls | Boys | Girls |
| 1. Who is strongest | 3.1 | 2.9 | 96.9 | 97.1 | 3.1 | 5.8 | 93.8 | 94.2 |
| 2. Who gives presents | 78.1 | 57.1 | 15.7 | 37.1 | 60.6 | 45.7 | 27.0 | 20.0 |
| 3. Who spanks most | 37.5 | 28.5 | 59.4 | 65.7 | 27.0 | 37.1 | 53.1 | 51.3 |
| 4. Who is boss | 9.3 | 25.7 | 87.6 | 71.4 | 15.7 | 14.3 | 56.2 | 68.6 |
| 5. Who child likes | 47.1 | 48.7 | 47.1 | 40.0 | 31.4 | 62.8 | 40.6 | 22.8 |
| 6. Who gets mad | 40.6 | 25.7 | 53.1 | 71.4 | 47.1 | 40.0 | 50.2 | 51.3 |
| 7. Who is smartest | 9.3 | 37.1 | 84.5 | 62.8 | 18.6 | 28.5 | 75.2 | 59.8 |
| 8. Who punishes | 37.5 | 25.7 | 59.4 | 71.4 | 27.0 | 25.7 | 56.2 | 45.7 |
| 9. Who is nicer | 60.6 | 68.6 | 37.5 | 31.4 | 56.2 | 59.8 | 31.4 | 31.4 |
| 10. Who child is scared of | 31.2 | 11.4 | 65.6 | 85.6 | 31.2 | 20.0 | 31.2 | 57.1 |
| 11. Who tells child what to do | 43.7 | 54.3 | 50.2 | 45.7 | 40.6 | 54.3 | 40.6 | 37.1 |
| 12. Who kisses child most | 78.1 | 42.9 | 21.9 | 54.3 | 75.2 | 51.3 | 15.7 | 34.3 |
| 13. Who will child grow up to be | 3.1 | 88.1 | 90.6 | 5.7 | 3.1 | 82.8 | 90.6 | 2.9 |

## RESULTS

### *Differential Perception of Mother and Father*

Table 1 presents the percentage of the boys and girls responding mother or father to the 13 questions in Set A (indirect method) and Set C (direct method).

Table 2 presents the percentage of boys and girls responding mother or father to the 10 pictures shown in Set B.

Initially, both boys and girls were divided into young vs. old in order to test for age differences on each of the three sets of questions. Since there were no statistically significant differences between the answers of the young and old children to any of the questions, the results presented are for all boys and all girls.

The data in Table 1 reveal that, except for Questions 5, 10, and 12, the answers to the indirect questions were quite similar to those obtained with the direct questions. For both the direct and indirect methods boys and girls saw the father, in relation to the mother, as stronger, the boss of the house, smarter, and the major agent of punishment (Questions 1, 3, 4, 7, and 8). To the question, "Who is the child more scared of?" boys and girls answered father more than mother on the indirect query. All of these differences were significant at the .05 level and all probability values presented are for two-tailed tests unless otherwise stated. The difference between mother and father answers to the direct question, "Who are you more scared of?" was not statistically significant for boys but was significant for girls ($p < .05$). That is, girls labeled father as more fear provoking than mother on both kinds of questions. To the indirect question, "Who gets mad?" the girls replied father more often than mother ($p < .05$). There was no difference for boys between father and mother answers on either type of question. Thus, girls labeled the father as more fear provoking and prone to anger than did the boys.

The mother, in relation to the father, was viewed as the giver of presents and nicer (Questions 2 and 9) on both methods ($p < .05$). To the question, "Who kisses the most?"

TABLE 2

PERCENTAGE OF BOYS AND GIRLS RESPONDING MOTHER OR FATHER TO THE TEN PICTURES

| Picture | Boys | | Girls | |
|---|---|---|---|---|
| | Mother | Father | Mother | Father |
| 1. Child given ice cream | 59.4 | 40.6 | 82.7 | 17.3 |
| 2. Hands reaching for toy | 25.0 | 75.0 | 11.4 | 88.6 |
| 3. Child being spanked | 40.6 | 56.3 | 31.4 | 68.6 |
| 4. Child picked up in air | 21.9 | 78.1 | 14.3 | 85.7 |
| 5. Child cowering | 12.5 | 87.5 | 22.8 | 77.2 |
| 6. Child given present | 87.5 | 12.5 | 68.6 | 31.4 |
| 7. Child running from dog | 28.1 | 71.9 | 31.4 | 68.6 |
| 8. Child taking walk | 53.1 | 46.9 | 51.4 | 45.6 |
| 9. Child on seesaw | 31.2 | 50.0 | 37.1 | 28.5 |
| 10. Child mad | 50.0 | 46.9 | 28.5 | 62.5 |

boys replied mother more often than father to both methods ($p < .05$). This difference did not hold for girls. The question, "Who tells the child what to do?" yielded no significant difference between the frequency of the mother and father answers.

The responses to the 10 pictures were consistent with these results. To Pictures 1 and 6, where a child was receiving a desirable object, both boys and girls responded mother more than father ($p < .05$). The pictures illustrating punishment (3 and 5) elicited more father responses ($p < .05$). The pictures depicting a child in a situation where he needs or receives assistance, strength, or protection (Pictures 2, 4, and 7) also elicited more father than mother answers ($p < .05$). There were no significant sex differences for any of the 10 pictures taken separately. The two pictures that were more neutral in content (8 and 9) did not produce significant differences between frequency of mother or father answers. Finally, it is of interest to note that to Picture 10, where the child was described as mad at someone, boys showed no significant difference between mother and father answers while girls responded father more than mother ($p < .05$). This result agrees with the answers to the indirect question: "Who gets mad at the child?" The girls replied father more often than mother while there was no difference in father-mother answers for the boys. It appears that with indirect procedures girls are more likely than boys to label the father as potentially hostile.

The results for all three methods agree with the research summarized earlier and reaffirm the statement that young children see mothers as more nurturant but less punitive, less fear arousing, and less competent than fathers.

*Method Differences*

There were only three questions that produced slightly different answer patterns in comparing the indirect with the direct method of questioning (Questions 5, 10, and 12). None of these differences were statistically significant for both sexes. To the question, "Who is the child more scared of?" the boys were more likely to answer father to the indirect than to the direct question ($p < .05$). To the question, "Who kisses most?" the girls answered father more to the indirect than to the direct question ($p < .05$). Thus, when the question dealt with fear of father boys showed the greatest shift in answers between the indirect and direct procedure. When the question dealt with physical affection girls showed the greatest shift between the two types of questions. To the question, "Who does the child like the best?" girls showed a small but nonsignificant increase in mother answers to the directly phrased question.

It is suggested that perhaps these three questions were more anxiety arousing than the remaining 10 and that the direct questioning produced sufficient anxiety to elicit some defensive processes. The direct question, "Who kisses you most?" may have elicited sexually tinged anxiety because of the implications of a physically affectionate relationship between the girl and her father. The fact that this shift did not occur for boys suggests that recognition or report of physical affection between a child and the opposite sex parent may be less anxiety arousing for boys than for girls.

The possibility that the direct method was more anxiety arousing finds support in the fact that the direct method elicited more evasive answers than the indirect method. Although the instructions emphasized that the child respond either mother or father, some children gave no response, answered both mother and father, or named another family figure like brother or sister. These "other" answers were much more frequent to the direct questions than to the indirect ones. The average number of children giving one or more "other" answers to the 13 indirect questions was 2.4 while the average was 10.4 for the direct questions ($p < .01$). About 15% of the children gave at least one "other" answer to the direct questions while only 3% of the group did so to the indirectly phrased questions. There were no sex differences in number of evasive replies.

*Sex Differences*

Sex differences did occur on some of the questions. The most striking sex difference occurred to the question: "Who would you like to grow up to be?" All children answered with the same sex parent on both indirect and direct methods ($p < .001$). The questions that dealt with affectionate and nurturant relations

between child and parent also produced sex differences. To the indirect question, "Who kisses most?" girls chose father and boys chose mother ($p < .01$). This difference did not hold for the direct question and it has been suggested that this might be due to the girl's anxiety over responding father to the direct query. To the direct question, "Who does the child like the best?" boys and girls tended to choose the same sex parent over the opposite sex parent ($p < .05$). Finally, the indirect question, "Who gives the most presents?" produced the slight tendency for girls to respond father more than boys ($p < .08$). The girls showed a tendency, on the indirect questions (Questions 2 and 12), to label the father as more nurturant than boys. However, the girls also showed a slight tendency on the indirect set to see the father as more punitive, more prone to anger, and more fear arousing than boys.

*Sex and Method Differences for Content Clusters*

Boys and girls differed in the degree of punitiveness and affection assigned to the father, despite the fact that all children agreed that mother was more nurturant and less punitive than father. In order to test the significance of this trend those questions that dealt with similar contents were considered as a cluster. The three clusters analyzed were punitiveness-hostility (Questions 3, 6, 8, and 10), nurturance-affection (Questions 2, 9, and 12), and competence-power (Questions 1, 4, and 7). The interitem correlations among the indirect questions in each of the three clusters were all positive and justified the pooling of these items. A child's score for father or mother answers on the punitiveness-hostility cluster ranged from 0 to 4; for the nurturance-affection and competence-power clusters the range was 0 to 3. The distributions of these scores for father and mother answers for each set of questions were highly skewed. Each distribution was divided at the median and chi square tests (with Yates' correction) were applied to assess sex and method differences for each of the three clusters. Table 3 presents the results of this analysis.

*Sex differences.* The direct method produced no significant sex differences. On the indirect method, the girls, in comparison to boys, characterized father as more punitive, more nurturant, but less competent ($p < .10$, .01, .01, respectively). The boys ascribed more competence to father, while girls attributed more competence to mother. This sex difference did not appear on the direct questions, suggesting that the indirect questions may be reflecting the child's wishes. Since there should be minimal anxiety over ascription of intelligence, strength, and power to a parent the fact that boys and girls show a decrease in same sex parent answers to the direct questions raises the possibility that the indirect answers were influenced by the child's desire to view the same sex parent as the more competent one.

*Method differences.* Boys were more consistent than girls across the two methods. To the direct questions, girls ascribed less punitiveness and less affection to father than they did on the indirect set. The boys' answers on

TABLE 3

SEX AND METHOD DIFFERENCES FOR THE PUNITIVE, NURTURANT, AND COMPETENCE CLUSTERS

| Cluster | Parent | Sex Differences | | Method Differences | |
|---|---|---|---|---|---|
| | | Indirect | Direct | Boys | Girls |
| Punitive | Father | G > B* | ns* | ns | I > D*** |
| | Mother | ns | ns | ns | ns |
| Nurturant | Father | G > B*** | ns | ns | I > D** |
| | Mother | ns | ns | ns | ns |
| Competent | Father | B > G*** | ns | I > D*** | ns |
| | Mother | G > B*** | ns | ns | I > D* |

Note.—B = Boys, G = Girls, D = Direct method, I = Indirect method.

* ns = not significant; $p > .15$; two tails.

* $p < .10$; two-tailed test.

** $p < .05$; two-tailed test.

*** $p < .01$; two-tailed test.

these two clusters were not significantly different across the two methods. The phi coefficients evaluating the reliability of father answers across the two methods were significant for boys at the .05 level for all three clusters (phi = .38, .39, and .29 for the punitive, nurturant, and competence dimensions, respectively). For the girls, the corresponding coefficients were .17, .00, and .41, the latter value significant at the .05 level.

Punitive and nurturant clusters of items were also selected from the 10 pictures in Set B. These answers were correlated with the responses to the punitive and nurturant clusters on the indirect and direct questions. Pictures 1, 4, and 6 were used for the nurturant cluster and Pictures 3, 5, and 12 for the punitive-hostile cluster. The relationship between answers to pictures and questions was assessed by phi coefficients. The answers to the indirect, punitive, and nurturant questions were significantly correlated with the answers to the punitive and nurturant pictures. That is, those children who labeled the father as punitive-hostile on the indirect questions also labeled the father as punitive-hostile on Picture Cluster 3, 5, and 12 (phi = .38; $p < .01$). Those children who labeled the mother as nurturant on the indirect question cluster replied mother to Picture Cluster 1, 4, and 6 (phi = .33; $p < .01$). These results held for both sexes. The answers to the direct, punitive, and nurturant clusters were not significantly related to the answers on the punitive and nurturant pictures.

## DISCUSSION

These results are in remarkably strong agreement with those of other studies (Emmerich, 1959; Finch, 1955; Gardner, 1947; Kagan, 1956) and indicate that children from preschool to preadolescence label the mother as more nurturant than the father. Father is labeled as more punitive, more competent, and a more frequent source of power and fear. Although all children agree that father is punitive and mother nurturant, the girls view the father with more ambivalence than the boys. That is, the girls label the father as both more hostile and more affectionate than the boys. It is suggested that the 6-year-old boy has begun to identify with his father, and is forced to repress some of the father's fear arousing and undesirable traits. The young boy, therefore, conceptualizes the father in a warmer and more benign light. The girl sees the father both as a more distant figure (because of the lack of identification with him) and as a potential love object and source of affection. This *post hoc* explanation would account for the girls' ambivalent, conscious perception of the father.

The data support Parsons' hypothesis that the mother role is predominantly nurturant while the paternal role is characterized by instrumental skills and competence. The results also agree with Brim's (1957) suggestion that each parent forms a different type of social relationship with a son or a daughter, e.g., the father-daughter relation is a different social system than the father-son interaction.

The authors do not believe that the child's perception of the parents is solely a function of each parent's behavior toward the child. The mass media, peer communications, and the content of children's books, all paint the adult male as competent, aggressive, and powerful. These sources of communication describe the mother as loving, weak, and subordinate (Bennett & Cohen, 1959; Child, Potter, & Levine, 1946). It is likely that the child's answers to our questions were influenced by these stereotyped conceptions of the mother and father roles. At a more speculative level, it is suggested that these extrafamilial communications also influence the child's perception of his parents.

That is, if a 5-year-old views men as potentially more aggressive and hostile than women there is reason to believe that he will generalize this expectation to his own father and view the latter as potentially more aggressive than he may be in reality. The child's conceptualizations of the significant roles and people in his life are apt to be the result of an interaction between his direct experiences with these social entities and the stereotyped labels he acquires through contact with symbolic representations of these roles.

*The acquisition of meaning and perception of parents.* The differential perceptions of father and mother seem congruent with the meaning factors described by Osgood et al. (1957). The reader will recall that Osgood found that the

adjectives large, strong, rough, hard, and heavy, which suggest power and masculinity, were positively intercorrelated. Each of these words showed zero or negative correlations with adjectives like kind, pleasant, and nice, which imply nurturance and femininity. It is suggested that the latter group of adjectives is more congruent with the child's conception of mother while the former is more characteristic of the concept of father. The present data suggest that the conceptual associations of adults, with respect to the attributes strength, power, and nurturance, have historical roots in the child's perception of his parents. Osgood reports that the word "nice" is closer in the semantic space to femaleness than it is to maleness. In the present study, the question, "Who is nicer to the child?" elicited more mother than father answers from both boys and girls. Thus, the purely conceptual association between the words "nice" and "female" discovered in college students is present in the conceptualization of mother and father by preschool children.

As mentioned earlier, Jenkins and Russell (1958) found that college students rated father, in relation to mother, as higher on the traits cruel, successful, strong, and bad. The current sample of children differentially labeled their parents in a similar way. This similarity suggests that such language learning begins very early in the child's conceptual development.

*Identification processes.* In an earlier paper on identification (Kagan, 1958) it was stated that the model the child would most likely choose for identification was the one who was viewed as competent and in control of sources of gratification. The present data suggest that girls view the father as a more competent and powerful figure than the mother. However, to the question, "Who do you want to be like when you grow up?" the girls answered mother. Thus, the role model that the girls have chosen is not the one whom they view as the most competent. It is tempting to speculate on the implications of this finding. Since the girl's identification influences both her conception of herself and her behavioral tendencies with other people, it is possible that the passivity of women with men is based, in part, on the child's belief that the father is stronger and more competent than the mother. These speculations are in essential agreement with other research on the sexual preferences of children. Brown (1956) asked 5- and 6-year-old children to select preferred activities and role behaviors. Although children typically preferred their own sex role, twice as many girls as boys indicated a preference for both male and female activities and roles. Moreover, boys showed a significantly greater preference for their own sex role than girls showed for the female role. On choosing between wanting to be a boy or a girl, only 10% of the boys said they desired to be a girl, while about a third of the girls preferred the male role. Smith (1939) has reported that, with age, both sexes come to have a better opinion of boys and progressively poorer opinion of girls during the ages 8 to 15. Smith asked 200 children whether boys or girls possessed a greater degree of desirable vs. undesirable traits. By age 14, girls felt boys were as high on the desirable attributes as were the girls.

The data suggest that girls have a more anxiety arousing identification and a less positive self-image than boys and that this phenomenon is influenced by the differential perception of fathers and mothers during early childhood.

*Methodological considerations.* This investigation suggests two tentative statements about the differential validity of indirect and direct methods of questioning with young children. Although there was a moderate degree of consistency across both procedures, the direct method elicited more evasive answers. The greatest differences between the two procedures occurred on those questions that might be regarded as anxiety arousing. One of the advantages of using both direct and indirect methods in a diagnostic or research setting is to obtain clues as to sources of anxiety. Lack of consistency between the answers to the two methods may indicate some anxiety associated with the content of the question.

Since the indirect questions elicited less evasiveness one might conclude that this technique gave a more accurate appraisal of the child's perceptions. The responses to the competence and power questions tend to contraindicate this conclusion. The children were more likely to see the same sex parent as smart and the boss on the indirect than on the direct questions. Since these questions should

not elicit high anxiety it is possible that the child's wish to view the same sex parent as more competent influenced his answers. Each method has unique characteristics and the combined use of the methods in research and diagnosis is perhaps the most sensible strategy.

## SUMMARY

A group of 67 children were individually interviewed with three different methods to obtain information on perception of the parents with respect to nurturance, punitiveness, source of fear, and competence.

Indirect questions: The children were first shown line drawings of a mother, father, and child and asked 13 questions about this "make believe" family. The questions pertained to parental nurturance, competence, punitiveness, and source of fear: e.g., "Who gives the child the most presents?" "Who is the child more scared of?"

Picture method: The children were then shown 10 different pictures illustrating a child in situations that usually involved a parent, but the parent was not illustrated. The child was asked to name who was missing from the picture, the mother or the father.

Direct questions: Finally, the child was asked the same 13 questions asked in the indirect method (in a different order) but the questions were phrased in terms of the child and his own parents: e.g., "Who gives you the most presents, your mother or your father?" "Who are you more scared of, your mother or your father?"

The results suggested the following conclusions:

On all three methods the fathers were seen as more fear arousing, more competent, and more punitive than mothers. Mothers were seen as "nicer" and more likely to give presents. This result held for both sexes.

There were a few sex differences. Boys and girls chose the same sex parent as the model they wished to emulate and the parent they "liked the best." They chose the opposite sex parent as the one who "kissed the most." Although all children labeled the father as punitive and mother as nurturant, the girls, in comparison with the boys, labeled the father as both more punitive and more affectionate. The girls seemed to perceive the father with more ambivalence than the boys.

The direct questions elicited significantly more evasive answers than the indirect questions. It was suggested that the indirect questions were most likely to elicit material different from direct questioning when the content of the question was anxiety arousing or was related to a wish of the child.

Girls stated a desire to be like the mother but perceived the father as wiser and stronger. It was suggested that girls have an anxiety arousing identification for their identification model is perceived as the least competent of the two parents.

The results supported both previous research in this area and Parsons' hypothesis regarding the differentiation of parental roles. The developmental implication of these results for Osgood's meaning factors was discussed.

## REFERENCES

BENNETT, E. M., & COHEN, L. R. Men and women: Personality patterns and contrasts. *Genet. psychol. Monogr.*, 1959, **60**, 101–153.

BRIM, O. G., JR. The parent-child relation as a social system: I. Parent and child roles. *Child Developm.*, 1957, **28**, 343–364.

BROWN, D. G. Sex-role preference in young children. *Psychol. Monogr.*, 1956, **70**(14, Whole No. 421).

CHILD, I. L., POTTER, E. H., & LEVINE, ESTELLE, M. Children's textbooks and personality development: An exploration in the social psychology of education. *Psychol. Monogr.*, 1946, **60**(3, Whole No. 279).

EMMERICH, W. Young children's discriminations of parent and child roles. *Child Developm.*, 1959, **30**, 404–420.

FINCH, HELEN M. Young children's concept of parent roles. *J. home Econ.*, 1955, **47**, 99–103.

GARDNER, PEARL L. An analysis of children's attitudes toward fathers. *J. genet. Psychol.*, 1947, **70**, 3–28.

JENKINS, J. J., & RUSSELL, W. A. An atlas of semantic profiles for 360 words. *Amer. J. Psychol.*, 1958, **71**, 688–699.

KAGAN, J. The child's perception of the parent. *J. abnorm. soc. Psychol.*, 1956, **53**, 257–258.

KAGAN, J. The concept of identification. *Psychol. Rev.*, 1958, **65**, 296–305.

MACCOBY, ELEANOR E. Role taking in childhood and its consequences for social learning. *Child Developm.*, 1959, **30**, 239–252.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana: Univer. Illinois Press, 1957.

PARSONS, T., & BALES, R. F. *Family, socialization and interaction process*. Glencoe, Illinois: Free Press, 1955.

SMITH, S. Age and sex differences in children's opinion concerning sex differences. *J. genet. Psychol.*, 1939, **54**, 17–25.

(Received September 28, 1959)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 448–456

# EFFECTS OF WHITE AUTHORITARIANISM IN BIRACIAL WORK GROUPS[1]

IRWIN KATZ AND LAWRENCE BENJAMIN
*New York University*

ALTHOUGH the California F Scale was originally developed to study the prejudiced personality, there have as yet been no investigations of the relationship between F Scale scores and actual behavior toward minority group members. It is not known, for example, how the social behavior of authoritarian whites differs from that of nonauthoritarians in various types of face-to-face contact with Negroes. Nor is it known how Negroes react to whites who differ in this regard. The present experiment attempted to study social interaction in work groups composed of two Negro men, and two white men with either high or low F Scale scores, under various conditions of reward and prestige.

Two lines of previous research on the F Scale are pertinent to this problem. First, it has been established that authoritarians are strongly prejudiced against cultural minorities. Adorno, Frenkel-Brunswik, Levinson, and Sanford (1950) reported positive correlations between the E (Ethnocentrism) and F Scales. Subsequent research, according to Christie and Cook (1958), supports this demonstrated relationship between the F Scale and prejudice. Second, Thibaut and Riecken (1955) found that authoritarians were more openly hostile against persons of low status than of high status, while variations in the status of the instigating person produced little difference in the expression of hostility by nonauthoritarians.

Thus, in the absence of external restraints one would expect authoritarian whites, as compared with nonauthoritarians, to aggress more against members of a low-status cultural minority, e.g., Negroes. But how do authoritarians behave toward Negroes when openly hostile acts are likely to bring forth retaliation, or punishment by an authority? Contact situations in which whites and Negroes have roughly equal status and in which there exists an official policy of nondiscrimination are becoming increasingly common in American life, both military and civilian. The present study was designed to observe interaction in a biracial situation of this kind. All men had identical job roles, received identical pay, and were treated alike in all respects by a white supervisor.

In addition to authoritarianism we were interested in two variables that gave promise of interacting with it, with both practical and theoretical implications: type of reward (group vs. individual) and level of group prestige.

*Reward.* Under a group bonus system, as compared with individual, noncompetitive bonuses, Negroes would have greater ability to retaliate against white unfriendliness, since by not cooperating on tasks they could prevent the latter from earning extra money. On the other hand, if group efforts were successful white men would tend to see Negro group members as agents of satisfaction, and consequently would be inclined to react more favorably to them. The first factor would inhibit anti-Negro behavior and the second would reduce instigations to such behavior. On the assumption that authoritarians are basically more hostile toward Negroes than are nonauthoritarians these two factors might be expected to produce greater differences in the behavior of authoritarian subjects (*S*s) than of nonauthoritarian *S*s.

*Prestige.* It is a reasonable assumption that in a condition of high group prestige, as compared with a neutral condition, white *S*s would tend to see Negro *S*s more as fellow members of a small élite and less as representatives of a disliked cultural minority, with the result that their behavior toward Negro *S*s would be more acceptant. On the supposition that authoritarians react to people primarily on the basis of perceived status, we expected that their social behavior would be more strongly affected by variations in group prestige than would the social behavior of nonauthoritarians.

[1] This is the second report on research carried out under Contract Nonr-285(24) between the Office of Naval Research and New York University. The project is under the direction of Irwin Katz. Judith March was a behavioral observer in the present experiment, in addition to the junior author.

*Racial differences for all conditions combined.* In an experiment on biracial groups Katz, Goldston, and Benjamin (1958) found that white men were more active, communicative, and productive than their Negro companions, and tended to talk mainly to one another, while Negroes addressed most of their remarks to whites and largely ignored one another. In the present study we wished to ascertain whether similar differences would occur on a wider range of measures when Negro and white group members were more closely matched on intelligence, and were repeatedly made aware of their mental equality, even to the point of being given tasks which were rigged to insure similarity of performance.

## METHOD

Negro and white men college students were employed to work in groups of four—two Negroes and two whites—on a series of tasks that were described as "materials that are being developed for vocational aptitude tests." Each group worked a total of 12½ hours, divided into four or five sessions, under a particular combination of white authoritarianism or white nonauthoritarianism, high prestige or neutral prestige, and group reward or individual reward. A 2 × 2 × 2 factorial design was used, with two groups in each cell. As dependent variables, social interaction, social perception, social influence, and productivity were measured by means of observers' classifications of ongoing behavior, questionnaires, and objective measures of output.

*Selection of Ss and matching on intelligence.* All *Ss* were from colleges and universities in New York City and between the ages of 19 and 30. Fifty-six Negroes and 160 whites were given the Verbal Reasoning and the Language Usage Tests of the Differential Aptitude Tests. The Verbal Reasoning Test became the primary instrument for matching men on intelligence, with scores on Language Usage being used to decide borderline cases.[2] The 32 highest scoring Negroes were selected from the original Negro pool of 56 men and matched as closely as possible with 32 whites (of whom 16 were identified as authoritarians and 16 as nonauthoritarians, by a method to be described below). Then, Negro-white matched pairs were randomly assigned to experimental groups. The mean Verbal Reasoning score of each race was 33, which corresponds to Percentile 65 in a national sample of high school seniors (Bennett, Seashore, & Wesman, 1947).

In the Negro sample the age range was 19 to 30 and the mean was 23.3. With one exception all Negro *Ss* had grown up in the North. In the white sample the age range was 19 to 28 and the mean was 21.2. Twenty-two white *Ss* were Jewish, five were Catholic, and five were Protestant.

*Establishment of the experimental conditions.* The laboratory consisted of a behavior room and an adjacent observational room connected by a one-way window and a one-way sound system. All instructions were given orally by the experimenter (*E*). Each group was told that it was to try out various potential test materials. *Ss* were also informed that they had been placed together because of the similarity of their DAT scores.

At the time of taking the DAT all men were administered a slightly modified version of the Webster, Sanford, and Freedman F Scale (1955).[3] Those in about the top third of the total white distribution of 160 scores were considered as high scorers and those in about the bottom third as low scorers. Sixteen high scorers and 16 low scorers were matched with Negro *Ss* on Verbal Reasoning scores and then assigned to work groups so that eight groups had 2 authoritarian whites and eight groups had 2 nonauthoritarian whites. The mean Verbal Reasoning scores of both Negro and white *Ss* in authoritarian groups and in nonauthoritarian groups were equated.[4]

All men were told that a bonus of 35¢ an hour, in addition to the base pay of $1.50 an hour, would be awarded for "effort." For induction of group reward *Ss* in eight groups were told that the bonus would be based on the effort of the group as a whole. For individual reward *Ss* in eight groups were told that the bonus would be awarded to each man individually on the basis of his own effort. It was stressed that the individual bonus was noncompetitive.

In eight high prestige groups men were told at the first session that on the basis of their DAT scores they were expected to perform better than other groups. At the start of the second and all subsequent sessions high prestige groups were shown graphs which favorably compared their scores on various tasks with those of 15 fictitious groups. Eight neutral prestige groups re-

[2] It might be objected that the DAT was designed for use with high school samples, and would have proved too easy for college students. However, this was not the case for our Negro *Ss* with whom white *Ss* were matched.

[3] We omitted 9 inappropriate items (all those under *D.3* "preference for traditional feminine role") from this 149-item scale that had been developed on a sample of Vassar undergraduates. It was used in preference to the original California scale because of its greater length and greater concentration on personality as opposed to ideology. In a sample of 100 white New York University men undergraduates, the Vassar scale showed a correlation of .79 with the California scale. In another sample of 67 from the same population its correlation with the Christie, Havel, and Seidenberg version of the F Scale (1958) was .69, and .26 with these authors' "acquiescence score." The latter correlation indicates that acquiescent response set was a factor of only minor importance in the Vassar F Scale scores of NYU men.

[4] For the total white distribution of 160 scores on the modified Vassar scale, $M = 540$ and $SD = 65$. Cutoff points for high scorers and low scorers were 580 and 497, respectively. For the total Negro pool of 56 men, $M = 539$ and $SD = 69$. Mean Negro F Scale scores for different experimental conditions ranged from 532 to 540.

ceived no information about their expected performance or actual performance relative to other groups.

*The tasks.* *Ss* worked on two 50-minute construction tasks with Erector Set materials, in which all had to build identical models from a picture. They were permitted to help one another.

Two human relations problems were discussed, each for 30 minutes. Groups had to agree unanimously about suitable answers to a set of questions. Before each discussion *Ss* answered the same questions privately.

A maze problem was rigged so that the average scores of Negroes and whites in the same group would be the same. The apparatus consisted of a 3 × 3 foot vertical wooden panel covered with bolt heads spaced 1 inch apart, and a metal stylus. The board was wired electrically so that when a "correct" bolt head was touched with the stylus a bulb lighted. A vertical path had to be learned. *Ss* worked individually, taking one trial at a time in rotating order. The number of trials required for an errorless run was secretly controlled from the observation room.

The maze apparatus contained two more complex paths which the men worked on as a team. Before each move *Ss* registered their opinions privately, after which they were required to reach a unanimous group decision through discussion. The latter determined the actual move.

A set of 20 multiple-choice abstract reasoning problems was rigged so that the average scores of Negroes and whites in a given group would be the same. After each problem the men announced their answers, and then the "correct" answer was announced by *E*.

Every group worked as a team for a maximum of 40 minutes on the Hanfmann-Kasanin Concept Formation Test.

Pairs of men worked on a two-man version of the ball-and-spiral game that was used by French (1953) and others. An open spiral tract 2½ inches wide was mounted on a 25 × 25 inch base, at each corner of which was a handle. Two men faced one another, each grasping two handles, and tried to maneuver a rubber ball up the tract and into a hole at the top. Every *S* worked ½ hour with the group member of his own race and ½ hour with a randomly selected group member of the other race. The order of pairing was counterbalanced for the various groups.

The order of tasks was as follows: Individual Maze, Construction 1, Group Maze 1, Abstract Reasoning, Human Relations 1, Construction 2, Hanfmann-Kasanin, Human Relations 2, Group Maze 2, Spiral.

*Dependent variables.* Communications were classified as they occurred by one of two observers, who were ignorant of the experimental treatments accorded particular groups. Twelve categories were used, as presented in Table 1 below. (Some were used on more than one task.) Interobserver correlations (rho) on number of communications initiated in various groups were computed separately for Negroes and whites. They ranged from .72 to .96 for different categories, with a median of .84. Six categories yielded, in addition to frequency scores, measures of the social direction of communications. For this purpose we employed an "S score" (Katz et al., 1958), which has to do with the proportion of remarks directed by an individual to own-race and to other-race recipients. On these scores the range of interobserver correlations was .65 to .92, with a median of .81.

Measures of opinion change under group influence were obtained on the group mazes and the human relations problems.

A terminal questionnaire included items on social preference, personality ratings, and evaluations of performance. Satisfaction with the group experience was measured with a 15-item Likert-type scale developed by Gibb (1954). Defensiveness toward the group was measured with a 25-item Likert-type scale reported by Smith (1956).

## RESULTS AND DISCUSSION

### *Main Effects of Authoritarianism on Communication*

In evaluating the effects of the independent variables on communication, the various categories of behavior were treated separately for each type of task, yielding a total of 21 measures of amount of communication for each race.[5] Each set of behavior scores was submitted to analysis of variance, taking into account three major variables—authoritarianism, reward, and prestige. We will report the main effects of authoritarianism only, but all interactions.[6]

*White.* Table 1 shows *F* ratios for main effects of the authoritarianism variable on number of communications initiated by white *Ss*, as well as interactions of authoritarianism with reward and prestige. Four main effects had probabilities below .10, including two significant main effects ($p < .05$) without interactions. Both of the latter occurred on spiral-mixed pairs: authoritarian whites accepted more action proposals from their Negro partners and rejected fewer than did nonauthoritarians. It is not surprising that the clearest behavioral differences between the two types of white *Ss* occurred when they were paired with Negroes on the spiral, for this task was intended to maximize interdependence by requiring close motor coordination between partners.

*Negro.* Table 2 shows all *F* values for main

[5] This does not include the group mazes, which were omitted because of a procedural error that resulted in time samples of indeterminate length.

[6] Similar analyses were performed on scores indicating direction of communication (S scores), but the effects of the experiment variables were so few and inconsistent that they are not included in this report.

TABLE 1
VALUES OF *F* FOR MAIN EFFECTS OF AUTHORITARIANISM AND INTERACTIONS WITH PRESTIGE AND REWARD ON NUMBER OF COMMUNICATIONS INITIATED BY WHITES

| Task and Category | Main Effects | | Interaction | | |
|---|---|---|---|---|---|
| | Author. Higher | Nonauthor. Higher | A/P | A/R | A/P/R |
| Construction | | | | | |
| Asks for help | 5.73** | | 10.52** | .68 | 12.60*** |
| Gives help | .45 | | .40 | .52 | .10 |
| Orients group toward gnal | | 1.20 | 1.19 | 1.01 | 9.57** |
| Human Relations | | | | | |
| Disagrees | 2.72 | | .97 | .64 | .39 |
| Agrees | .90 | | .53 | .48 | 17.10*** |
| Gives information or opinion | .45 | | 1.01 | .66 | 1.20 |
| Asks for info., etc. | .79 | | .89 | .54 | 1.42 |
| Hanfmann-Kasanin | | | | | |
| Disagrees | .37 | | .11 | .85 | .15 |
| Asks for opinion | | .32 | .46 | .28 | .48 |
| Gives opinion | | 1.05 | .32 | .43 | .53 |
| Orients group | | 1.24 | .79 | .67 | .28 |
| Spiral—Mixed Pairs | | | | | |
| Accepts proposal | 16.15*** | | .61 | .47 | .55 |
| Proposes action | .88 | | 8.44** | .22 | 1.66 |
| Praises other | | .25 | 7.50** | .19 | .50 |
| Expresses hostility | | .36 | 1.62 | .59 | .73 |
| Rejects proposal | | 5.86** | .40 | .14 | .76 |
| Spiral—White Pairs | | | | | |
| Expresses hostility | 3.76* | | 8.43** | .93 | 1.16 |
| Proposes action | 1.08 | | 5.73** | .88 | .64 |
| Rejects proposal | .90 | | .34 | .28 | .78 |
| Accepts proposal | | 1.44 | .17 | 1.90 | 3.32 |
| Praises other | | 1.65 | 12.40** | 4.31* | .85 |

Note.—For all *F* ratios, within-cells variance was used as the error term, with 8 *df*.
* $p < .10$.
** $p < .05$.
*** $p < .01$.

effects of authoritarianism and its interactions with reward and prestige on number of communications initiated by Negroes. There were 10 main effects with probabilities below .10, including 7 without interactions (of which 4 were significant at $p < .05$). In general, the main effects depict Negroes in authoritarian groups as both more assertive and more cooperative than Negroes in nonauthoritarian groups. It appears that Negroes were more strongly affected in overt behavior by the F Scale scores of white *S*s than were whites themselves. Hence, we must assume that important behavioral cues were presented to Negro men by their white companions which our observational categories did not pick up.

### *Main Effects of Authoritarianism on Social Influence*

Table 3 presents values of *F* for main effects of authoritarianism and its interactions with reward and prestige on the amount of change in *S*s' opinions under group influence. For the two group mazes each *S*'s change score was the total number of times his private prediscussion opinions about moves differed from subsequent unanimous group decisions. His score on the two human relations tasks was the sum of differences between his private prediscussion position on an attitude scale and the position later agreed on unanimously by the group. Table 3 shows that on both types of task there were trends ($p < .10$) in which

TABLE 2
VALUES OF *F* FOR MAIN EFFECTS OF AUTHORITARIANISM AND INTERACTIONS WITH PRESTIGE AND REWARD ON NUMBER OF COMMUNICATIONS INITIATED BY NEGROES

| Task and Category | Main Effects | | Interaction | | |
|---|---|---|---|---|---|
| | Author. Higher | Non-author. Higher | A/P | A/R | A/P/R |
| Construction | | | | | |
| Gives help | 5.33** | | .12 | .93 | 2.90 |
| Asks for help | .47 | | .92 | 16.72*** | .63 |
| Orients group | .14 | | .61 | .50 | 1.08 |
| Human Relations | | | | | |
| Asks for info., etc. | 5.90** | | .19 | 1.54 | .55 |
| Disagrees | 5.71** | | .71 | .67 | .14 |
| Gives info., etc. | 5.25* | | .23 | .88 | .27 |
| Agrees | .66 | | 1.46 | 1.01 | .78 |
| Hanfmann-Kasanin | | | | | |
| Gives opinion | 14.07*** | | .98 | 15.05*** | 2.02 |
| Disagrees | 2.25 | | .16 | 8.72** | .38 |
| Orients group | 1.85 | | 1.65 | 2.10 | .62 |
| Asks for opinion | | .79 | .47 | 2.70 | .96 |
| Spiral—Mixed Pairs | | | | | |
| Proposes action | 21.63*** | | 1.75 | 6.93** | 3.87 |
| Praises other | 20.20*** | | .38 | 6.73** | 4.83* |
| Accepts proposal | 1.65 | | 3.32 | .64 | .18 |
| Expresses hostility | .40 | | .93 | .21 | .35 |
| Rejects proposal | | 4.92* | .81 | .40 | .25 |
| Spiral—Negro Pairs | | | | | |
| Proposes action | 10.91** | | .33 | 3.26 | 2.31 |
| Accepts proposal | 3.83* | | 1.82 | 1.82 | .69 |
| Praises other | .88 | | 2.24 | 1.84 | .20 |
| Expresses hostility | | 1.32 | .47 | .19 | .53 |
| Rejects proposal | | 2.00 | 2.61 | 2.61 | .96 |

* $p < .10$.
** $p < .05$.
*** $p < .01$.

TABLE 3
VALUES OF $F$ FOR MAIN EFFECTS OF AUTHORITARIANISM AND INTERACTIONS WITH PRESTIGE AND REWARD ON AMOUNT OF OPINION CHANGE DUE TO GROUP INFLUENCE

| | Main Effects | | Interaction | | |
|---|---|---|---|---|---|
| | Author. Higher | Nonauthor. Higher | A/P | A/R | A/P/R |
| Human Relations | | | | | |
| Negroes | .22 | | .88 | .53 | .26 |
| Whites | 4.70* | | .44 | .79 | 1.01 |
| Mazes | | | | | |
| Negroes | .09 | | .30 | 1.02 | .77 |
| Whites | 4.38* | | .19 | .56 | 1.82 |

* $p < .10$.

authoritarian white *Ss* shifted more than nonauthoritarian whites, while Negroes seemed to be unaffected by the experimental conditions.

*Main Effects of Authoritarianism on Personality Ratings*

Group members rated one another on 11 trait variables at the termination of their group experience. The complete list of trait variables appears in Table 4. Scores on each variable, as well as two combinations of scores, were submitted to analysis of variance tests. Table 4 shows values of $F$ for main effects of authoritarianism and its interactions with reward and prestige on ratings by white *Ss*. There were 2 significant main effects ($p < .05$) of authoritarianism on white *Ss'* ratings of one another, out of a total of 13 single and combined ratings. Both occurred without significant interactions. Authoritarian whites rated one another more "modest" and less "frank" than did nonauthoritarian whites. On white ratings of Negroes there were four significant main effects and one trend. Authoritarian whites rated Negroes higher on "bright," "mature," and "dominating" than did nonauthoritarian whites; they also gave Negroes a higher Total Ascendancy score, and tended to rate them higher on "leads." These results are consistent with findings presented above on the communicative behavior of Negro *Ss*.

Negro ratings of group members were not influenced by the experimental variables, except for two trends toward main effects of authoritarianism: Negroes rated nonauthoritarian whites as brighter ($F = 4.00$; $p < .10$) and less cooperative ($F = 3.55$; $p < .10$).

*Results on Productivity and Other Dependent Variables*

Productivity was measured for the group mazes (number of errors), construction tasks

TABLE 4
VALUES OF $F$ FOR MAIN EFFECTS OF AUTHORITARIANISM AND INTERACTIONS WITH PRESTIGE AND REWARD ON WHITE SUBJECTS' RATINGS OF OTHERS' TRAITS

| Item | Ratings of Whites | | | | Ratings of Negroes | | | |
|---|---|---|---|---|---|---|---|---|
| | Author. Higher | Nonauthor. Higher | A/P | A/R | Author. Higher | Nonauthor. Higher | A/P | A/R |
| Modest-conceited | 14.47*** | | | 2.85 | | .48 | .43 | .80 |
| Warm-cold | | .95 | 6.60** | .42 | | .93 | 4.16* | .64 |
| Bright-dull | | .55 | 3.77* | .12 | 8.75** | | .82 | .48 |
| Relaxed-tense | .31 | | .87 | .23 | 3.26 | | .19 | .16 |
| Good mixer-poor mixer | | .81 | 12.23*** | 3.31 | 1.85 | | .59 | .36 |
| Mature-immature | 1.11 | | 5.08* | .50 | 5.44** | | 4.00* | 9.00** |
| Frank-concealing | | 8.25** | 2.50 | .38 | | .29 | .91 | .20 |
| Cooperative-uncooperative | 1.27 | | 6.79** | 4.45* | 1.35 | | .35 | .13 |
| All Evaluative scales (above) | | .69 | 13.89*** | .30 | 1.59 | | 2.18 | .33 |
| Dominating-submissive | | 1.43 | .99 | .69 | 7.33** | | .84 | 7.33** |
| Leads-follows | | .60 | 5.11* | .51 | 4.33* | | .62 | .25 |
| Active-passive | | 1.13 | 8.00** | .87 | 2.82 | | .93 | .61 |
| All Ascendancy scales (above) | | 1.80 | 3.54* | .66 | 7.16** | | .45 | .45 |

Note.—Triple interactions are omitted. None were significant.
* $p < .10$.
** $p < .05$.
*** $p < .01$.

(accuracy of models), the Hanfmann-Kasanin problem (solution time), and the ball-and-spiral (number of times various levels were reached, weighted for relative difficulty of levels). On the spiral task pairs of nonauthoritarian whites achieved significantly higher scores than did pairs of authoritarian whites ($F = 9.76$; $p < .05$). To save space other $F$ ratios for productivity are not presented, since there were none that corresponded to $p < .10$.

Similarly, findings on the Defensiveness scale, the Satisfaction scale, social preference, and evaluations of performance are omitted because there were no values of $F$ at $p < .10$.

*Discussion of Main Effects*

Authoritarian whites tended to behave more compliantly in biracial situations and evaluated Negroes more favorably than did nonauthoritarian whites, while Negroes were more ascendant with authoritarians than with nonauthoritarians. Our results are understandable in terms of the low perceived power of authoritarians in the present situation. The unbiased behavior of the white *E*, the university setting of the experiment, and the numerical and intellectual equality of white and Negro *S*s may have introduced strong restraints against open expression of anti-Negro sentiments. The fact that authoritarians were actually more deferential with Negroes than were nonauthoritarians, suggests the presence of internal controls so powerful as to produce behavior opposite to their hostile impulses. In line with this interpretation, there was no evidence of greater cooperativeness or more favorable trait ratings *between* authoritarians than *between* nonauthoritarians. Indeed, authoritarians' ratings of one another as more "modest" and less "frank" suggest awareness of a cautious, tactical quality of behavior. And when paired together on the spiral, authoritarians were less able to cooperate than were nonauthoritarians, as shown in lower achievement scores.

*Interactions of Experimental Variables on Communication and Trait Ratings*

*White scores.* In Table 1 it can be seen that there were six significant interactions of authoritarianism and prestige on white communication frequencies. And in Table 4 there are five significant interactions and six trends to interaction of these two variables on white *S*s' descriptions of other group members. Between-cell comparisons by $t$ tests were not made because of the smallness of cell samples. However, inspection of cell means revealed perfectly consistent patterns in which scores on cooperative behavior and favorableness of personality ratings were always relatively high in the authoritarian-prestige treatment and in the nonauthoritarian-no prestige treatment, and were always relatively low in the other two treatments.

That authoritarian whites tended to respond more favorably to prestige than to no prestige is consistent with the general theory of the authoritarian personality. But it is unclear why nonauthoritarian *S*s tended to respond better to no prestige than to prestige. Possibly the ego-defensiveness of men in high prestige groups that was noted by Back (1951) and Katz et al. (1958) is especially strong in low F Scale scorers, who presumably are self-critical and would not readily accept an arbitrarily high appraisal of their ability (such as was given in this experiment by *E*).

A total of three significant triple interactions (authoritarianism × prestige × reward) on white communication scores had no consistent patterns of cell means.

*Negro scores.* There were five significant interactions of authoritarianism and reward on Negro communication frequencies (see Table 2). Inspection of cell means revealed that Negroes in groups with authoritarian whites were always more active under individual reward than under group reward, while Negroes in nonauthoritarian groups showed no consistent differences in behavior or ratings under the two reward conditions. We conjecture that Negroes perhaps were able to sense the presence of inhibited hostility in authoritarian whites, hence, tended to withdraw further from them in the condition of relatively high interdependence (group reward) than in the condition of relatively low interdependence (individual reward); but sensing little or no hostility in nonauthoritarian whites the Negro *S*s behaved about the same under both conditions of reward.

TABLE 5
VALUES OF $t$ FOR RACIAL DIFFERENCES IN AMOUNT OF COMMUNICATION

| Category | Task | | | | |
|---|---|---|---|---|---|
| | Construction | Maze | Human Relations | Blocks | Spiral[a] |
| Gives help | 1.79 | | | | |
| Asks for help | 3.52*** | | | | |
| Gives info., etc. | | 3.63*** | .65 | 1.58 | |
| Asks for info., etc. | | 1.87* | 1.46 | 1.50 | |
| Agrees | | .49 | 3.56*** | | |
| Disagrees | | 4.56*** | 4.45*** | 2.32** | |
| Proposes action | | | | | .10 |
| Accepts proposal | | | | | 1.95* |
| Rejects proposal | | | | | .23 |
| Orients group | 2.70** | | | .91 | |
| Praises other | | | | | 1.99* |
| Expresses hostility | | | | | 1.58 |

Note.—For all values of $t$ above 1.0 white scores were higher than Negro scores.

[a] Negro-white pairs only.

* $p < .10$.

** $p < .05$.

*** $p < .01$.

*Racial Differences for all Conditions Combined*

Katz et al. (1958) found that for all experimental conditions combined, white *Ss*, who were unselected with regard to F Scale scores, proved to be more active, communicative, and productive than their Negro companions, and tended to talk mainly to one another, while Negro *Ss* oriented compliantly toward the whites. In the present study we wanted to see whether racial differences similar to those obtained previously would occur—this time on a wider range of dependent measures—when Negro and white *Ss* were closely matched on intelligence and displayed equal mental competence (in terms of contrived scores).

To compare Negroes and whites on a given dependent variable a difference score was computed for each experimental group by subtracting the combined scores of the two Negroes from the combined scores of the two whites. For each distribution of difference scores a $t$ test was used to evaluate the divergence of the mean from a theoretical value of zero. Since the 16 experimental groups had been subjected to eight experimental treatments (two in each treatment) the estimate of variance had 8 *df*.

*Amount of communication.* Table 5 shows, for each of 20 observational measures, the $t$ value of the difference in number of communications initiated by Negroes and by whites. (Maze data are here included, since the sampling irregularities mentioned in the previous section did not invalidate Negro-white different scores.) Three trends and seven significant findings occurred, in all of which white scores were higher than Negro scores. Thus, on half of all the communication measures Negroes appeared to be less active than whites.

*Direction of communication.* S scores, as described by Katz et al. (1958), were used to measure the proportion of communications of various types that were directed to own-race persons and to other-race persons. These scores were not applied to spiral behavior, since in that situation only two men were present at a time. Two additional categories could not be scored for direction because of zero frequencies in some groups. In Table 6 are presented values of $t$ for directional "bias"—i.e., deviations from random targeting of remarks—on the part of white and Negro initiators. With respect to white initiators, out of 13 scoreable categories, there were 5 on which the probability that targeting was random was less than .10. On 3 categories white *Ss* favored one another as recipients, while on 2 categories they favored Negroes as recipients.

Table 7 shows that Negroes, on the other hand, spoke mainly to whites on 11 out of 13 categories (two trends and nine significant differences) and favored one another on none.

*Social influence.*[7] On the group mazes Negroes shifted more than whites from their private choice-point decisions to subsequent group decisions ($t = 2.44$; $p < .05$). The greater compliance of Negroes was not due to differences in maze learning ability, for there were no reliable differences in number of errors made by Negroes and by whites in their private decisions.

*Social preference.* On one of three sociometric items—preference as future working companions—Negroes favored one another ($t = 2.5$; $p < .05$), while whites showed no bias.

*Rankings on task competence.* At the end of the contact experience *Ss* were required to

[7] For measures of social influence and all remaining dependent variables only results at $p < .10$ will be presented, in order to save space.

TABLE 6
VALUES OF $t$ FOR TESTS OF "BIAS" IN DIRECTION OF WHITE COMMUNICATIONS

| Category | Task and Favored Recipients | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Construction | | Maze | | Human Relations | | Blocks | |
| | White | Negro | White | Negro | White | Negro | White | Negro |
| Gives help | 3.19** | | | | | | | |
| Asks for help | 2.41** | | | | | | | |
| Gives info., etc. | | | .25 | | .22 | | .43 | |
| Asks for info., etc. | | | | 1.09 | | 3.83*** | 3.56*** | |
| Agrees | | | | 1.74 | | 2.07* | | |
| Disagrees | | | 1.28 | | | 1.67 | | 1.05 |

* $p < .10$.
** $p < .05$.
*** $p < .01$.

rank one another for competence on the tasks they had performed. In analyzing these data we combined rankings on mazes, Hanfmann-Kasanin, and abstract reasoning into a single "cognitive tasks" score, while construction, spiral, and human relations were treated separately. Negroes ranked whites higher than they ranked each other on "cognitive tasks" ($t = 3.27$; $p < .05$) even though their objective performance was virtually the same as that of whites on the mazes and abstract reasoning and not reliably different on Hanfmann-Kasanin.

*Satisfaction.* Whites scored higher than Negroes on the Group Satisfaction scale ($t = 2.76$; $p < .05$).

Thus, even though Negro *Ss* were closely matched with white group mates on intelligence and were given evidence of equal ability on cognitive tasks they maintained a compliant orientation toward the white men and evidenced relatively low satisfaction and self-confidence. White *Ss*, on the other hand, gave no measurable indications of anti-Negro bias either in overt behavior or questionnaire responses.

TABLE 7
VALUES OF $t$ FOR TESTS OF "BIAS" IN DIRECTION OF NEGRO COMMUNICATIONS

| Category | Task[a] | | | |
|---|---|---|---|---|
| | Construction | Maze | Human Relations | Blocks |
| Gives help | .72 | | | |
| Asks for help | .64 | | | |
| Gives info., etc. | | 4.67*** | 5.04*** | 2.69** |
| Asks for info., etc. | | 2.29* | 4.87*** | 2.27* |
| Agrees | | 4.43*** | 2.61** | |
| Disagrees | | 6.36*** | 3.09** | 7.74*** |

[a] White recipients were always favored.
* $p < .10$.
** $p < .05$.
*** $p < .01$.

## SUMMARY

Sixteen groups consisting of two Negro and two white college students who were matched on intelligence worked under different combinations of group or individual reward and high or neutral group prestige. In half of the groups both white *Ss* had high F Scale scores and in the other half both white *Ss* had low F Scale scores. Measures of communication, social influence, productivity, and social perception were obtained.

Comparing authoritarian whites and non-authoritarian whites: on a task that required close cooperation between two *Ss*, authoritarians accepted significantly more suggestions from Negro *Ss* and rejected significantly fewer; on two tasks requiring group decisions authoritarians showed trends toward greater compliance; authoritarians rated Negroes significantly higher on intelligence, maturity, and dominance; authoritarians were significantly less productive when paired together on a two-man task requiring cooperation. The first three effects were interpreted as due to the authoritarian's fear of revealing anti-Negro attitudes in a potentially punitive environment.

Negroes who worked with authoritarians

were generally more assertive and cooperative than Negroes who worked with nonauthoritarians, as indicated by trends or significant differences on 10 out of 21 communication categories.

For white *Ss* there were trends or significant interactions of authoritarianism and prestige on 6 out of 21 types of communication frequencies and on 11 out of 26 ratings of others' traits. In all of these interactions inspection of cell means revealed that behavior was more cooperative or ratings were more favorable in the authoritarian-prestige and nonauthoritarian-no prestige treatments than in the nonauthoritarian-prestige and authoritarian-no prestige treatments.

For Negroes there were significant interactions of authoritarianism and reward on five types of communication frequencies, and inspection showed that Negroes grouped with authoritarian whites were always more active under individual reward than under group reward, while for Negroes in nonauthoritarian groups there were no consistent differences in behavior under the two conditions of reward.

For the entire sample of 16 groups it was found significantly that whites spoke more than Negroes, Negroes spoke more to whites than to one another, Negroes were more susceptible to group influence than were whites, ranked whites higher on mental ability, favored one another as future work companions, and scored lower than whites on a Group Satisfaction scale. These results indicate that even when Negroes are given objective evidence of equal mental ability in a relatively brief interracial contact they tend to feel inadequate and to orient compliantly toward whites.

## REFERENCES


ADORNO, T. W., FRENKEL-BRUNSWIK, ELSA, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.

BACK, K. Influence through social communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 9–23.

BENNETT, G. K., SEASHORE, H. G., & WESMAN, A. G. *A manual for the Differential Aptitude Tests*. New York: Psychological Corp., 1947.

CHRISTIE, R., & COOK, PEGGY. A guide to published literature relating to the authoritarian personality through 1956. *J. Psychol.*, 1958, **45**, 171–199.

CHRISTIE, R., HAVEL, JOAN, & SEIDENBERG, B. Is the F Scale irreversible? *J. abnorm. soc. Psychol.*, 1958, **56**, 143–159.

FRENCH, J. R. P., JR. The disruption and cohesion of groups. In D. Cartwright & A. Zander (Eds.), *Group dynamics*. Evanston, Ill.: Row, Peterson, 1953.

GIBB, J. R. Factors producing defensive behavior within groups. Quarterly Status Report, Feb. 15, 1954, Contract Nonr-1147(03) NR 170-226.

KATZ, I., GOLDSTON, JUDITH, & BENJAMIN, L. Behavior and productivity in bi-racial work groups. *Hum. Relat.*, 1958, **11**, 123–141.

SMITH, E. E. Effects of threat induced by ambiguous role expectations on defensiveness and productivity in small groups. Technical Report No. 1, Aug. 1956, Contract Nonr-1147(03) NR 170–226.

THIBAUT, J., & RIECKEN, H. W. Authoritarianism, status, and the communication of aggression. *Hum. Relat.*, 1955, **8**, 95–120.

WEBSTER, H., SANFORD, N., & FREEDMAN, M. A new instrument for studying authoritarianism in personality. *J. Psychol.*, 1955, **40**, 73–84.


(Received September 28, 1959)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 457–462

# SIZE ESTIMATION OF PICTURES ASSOCIATED WITH SUCCESS AND FAILURE AS A FUNCTION OF MANIFEST ANXIETY[1]

THEODORE P. ZAHN
*National Institute of Mental Health*

An unsettled problem of considerable theoretical interest concerns the relationship of attitudinal and motivational variables to judgments of magnitude. In various studies perceptual accentuation of the sizes of objects has been found to be a function of positive monetary (Bruner & Goodman, 1947; Dukes & Bevan, 1952) and symbolic value (Bruner & Postman, 1948), negative monetary (Dukes & Bevan, 1952) and symbolic value (Bruner & Postman, 1948), and positive reinforcement (Lambert, Solomon, & Watson, 1949). Granting the soundness of these empirical findings, the generalization is tenable that the personal relevance or importance of the object or symbol judged was the important variable underlying all of the above experiments.

A method of increasing the personal relevance of an object might be to make that object central in a task on which the subject (*S*) fails. If negative reinforcements occur in close conjunction with responses to the object, the latter may acquire secondary reinforcing properties so that it will evoke the same kind of emotional reaction as is produced by the negative reinforcement itself. If the above considerations are true the size of the object so reinforced should be overestimated.

In a previous study (Zahn, 1959), it was found that certain schizophrenic patients (i.e., those with good premorbid histories) accentuated the sizes of a picture associated with failure and of a picture of a mother scolding a boy relative to more positively toned pictures. It was hypothesized there that the patients showing this effect were characterized by more affective sensitivity or anxiety than were the patients who did not show the effect and than the normal controls who gave equivocal results.

The present experiment investigates the size estimation, from immediate memory, of pictures that have been associated with success and failure in a prior task by *S*s with varying degrees of manifest anxiety. The size estimations of these two pictures are compared with each other and with estimations of the sizes of pictures which have *not* been involved in a prior task. It was hypothesized that the failure picture would be overestimated relative to the success and neutral pictures primarily by *S*s with a high degree of manifest anxiety.

## METHOD

The general procedure consisted of three parts: a session of size estimation during which the sizes of three pictures were judged, four trials per picture, by the method of average error; a discrimination task in which success and failure experiences were provided for two additional pictures; and another session of size estimation in which the two "treated" pictures and two others were judged as before.

One of the purposes of the experiment was to record psychophysiological measures (GSR and heart rate) during the procedure. Hence, the experiment was introduced as a study to determine how such measures behaved while *S*s were performing simple tasks. Four electrodes were secured to *S*'s left hand and arm and right leg during the experiment. The data thereby secured are not reported here, however.

*Subjects.* *S*s were 48 undergraduate students at Duke University. Of these, 29 "volunteered" for credit from sections of the spring semester course in elementary psychology and 13 from the summer session course. Six additional *S*s were paid for their participation. Ten of the *S*s were women.

The Taylor Manifest Anxiety Scale (MAS), plus 15 items from the MMPI Lie (*L*) scale, had been given to the psychology classes from which 42 *S*s were drawn. The six *S*s not taking psychology were administered the scale just before the experiment. The experimenter (*E*) was unaware of the scores of the *S*s at the time of the experiment. *S*s from the entire range of MAS scores were used. For most of the analyses in which two MAS groups were considered, the low anxious (LA) *S*s had scores of 4–13 and the high anxious (HA) *S*s 14–32. The results for three additional *S*s were discarded because of *L* scores of six or more.

*The size estimation task.* The apparatus for size estimation has been described in detail elsewhere (Harris, 1957). It allowed for the projection of an outline drawing (white lines on dark background) onto a translucent screen of tightly stretched draftsman's parchment large enough to minimize framework cues.

[1] This study was carried out at Duke University and supported by USPHS NIMH Grant M-629. The author wishes to express his sincere thanks to Eliot H. Rodnick for his suggestions and encouragement.

The size of the image on the screen could be adjusted by means of a knob protruding below the screen. The image remained in focus and constant in brightness at all sizes.

*S* estimated the sizes of the pictures by the method of adjustment. On each trial, the picture was exposed at the standard size for about 5 seconds by count; then it was shut off for about 4 seconds and re-exposed at a different size. *S* then adjusted the image to what he estimated to be the standard size. Four such trials were run on each picture.

The stimulus pictures were as follows: (*a*) a square; (*b*) two business buildings, one taller than the other (Bl); (*c*) a tree and a bush (Tr); (*d*) a house and a tree (HT); (*e*) a lamp and a table (LT); (*f*) a "feeding" picture of a "mother" about to pour a glass of milk out of a pitcher for her "son" who is holding a glass in his outstretched hand (Fe); and (*g*) a "scolding" picture of a "mother" with her arm raised in a scolding manner over a boy who stands with head down (Sc). *Ss* were asked to imagine the most unpleasant instance involving being scolded by their mothers that they could while judging this picture.

All pictures were simple line drawings. Just before *S* was shown the picture for either of the two tasks, *E* told him which picture he would see with essentially the same description as presented here. Bl and Tr are referred to as "control" pictures. HT and LT were the pictures associated with success and failure as described below.

The height of the tallest object in each picture at the standard size was 16.5 centimeters on each trial. The sizes (height) at which the pictures were re-exposed were 13.5, 18.9, 21.2, and 14.8 centimeters in that order, constant for all *Ss* and for all pictures judged.

*The discrimination task.* The discrimination task was used solely as a means for creating a success and a failure experience in association with different previously neutral pictures. The instructions stated that *E* was interested in how well *S* could tell the difference between two pictures when they were presented rapidly and that the crucial difference would be in some detail in the picture, *not* size or brightness.

The pictures were exposed by a Revere automatic projector, the gears of which were modified so as to allow for an exposure of approximately .20 second. On each trial the standard picture was exposed for about 2 seconds, then 2 seconds later the variable picture was flashed for .20 second. *S* indicated "same" or "different" by operating a Mallory switch which lit up either of two lights for *E*. No time requirements were put on this response. About 5 seconds after *S*'s response, *E* reinforced it by lighting up one of two boxes facing *S* just below the screen. The word "right" was illuminated in the aperture of one box in white letters; the word "wrong" in red letters was illuminated in the other box.

There were 30 such trials, 15 on each of two pictures. On the success picture, the standard and variable exposures differed in a small but discriminable way on 10 of the 15 trials; and *S*, regardless of his response, was flashed "right" on 13 of the 15 trials and "wrong" on 2 trials. On the failure picture, however, the standard and variable exposures were always *identical*; and *S* was flashed "wrong" on 9 or 13 of the trials, depending on the condition (see below), and "right" on the remainder.

The two pictures so reinforced were the HT and LT pictures. On variations of the HT picture, a limb of the tree varied in angle, and on the LT picture, a leg of the table varied in angle. These pictures were exposed at exactly the same size as the standard size of the size estimation procedure.

There were three conditions (for three groups of *Ss*), varying in the sequence of pictures or the number of "wrongs" and "rights." In Condition 1, trials on the two pictures were mixed in a fixed irregular order, and the last 10 trials of the failure picture were uniformly reinforced by "wrong." Condition 2 was identical to Condition 1 except that on the last 10 trials of the failure picture, 7 of *S*'s responses were reinforced by "wrong" and 3 by "right." In Condition 3, trials on the success and failure pictures were alternated in a regular ABAB order, and the reinforcement schedule was the same as in Condition 2. This procedure was undertaken mainly because of the presumed effects on the psychophysiological measures.

For each of these conditions, approximately half the *Ss* were successful on the HT picture and judged the size of it first; and half were successful on the LT picture and judged it second. These will be designated as orders S-First and F-First, respectively. Although this procedure confounds the effects of the order of judgment with the picture succeeded or failed on, it serves to counterbalance the effects of both.

*Procedure.* In the first session of size estimations, *S* judged the sizes of the square, either the Tr or Bl pictures (counterbalanced), and the Fe picture in that order. After the discrimination task, the two reinforced pictures (HT and LT) were judged, followed by whichever of Tr or Bl that was not previously judged. In order S-First, the Bl picture was judged in the first session and the Tr picture in the second, whereas in order F-First, the sequence of neutral pictures was the reverse.

*Treatment of data.* The data are expressed as the deviation of the adjusted size from the standard size in terms of the distance of the slide away from the light source in mm. This measure was used in an attempt to approximate the discriminability function for size. It very closely approximates a logarithmic function of the linear dimensions of the image on the screen and, hence, should be closely related to discriminable size differences. The score for any *S* on any one picture is the mean of four of these deviations.

## RESULTS

### *Size Estimation and Total MAS Scores*

Table 1 shows means and variances of the size estimation scores for the two anxiety groups on the two reinforced and nonreinforced pictures. All mean scores represent *underestimations* of the size of the picture compared to the standard size, the more negative the score, the greater the underestimation.

A trend for generally more underestimation in the HA group is indicated except in the case of the failure picture which is relatively overestimated by this group. This picture departs from the LA group trend in the direction of greater underestimation.

The most sensitive and least contaminated test of the hypothesis that success and failure experiences influenced the size estimations of the two anxiety groups differentially is to compare differences in this respect between the two pictures for each *S*. A preliminary analysis of these difference scores, which included the three conditions under which the discrimination task was administered, two orders of size estimation, and the two levels of anxiety as independent variables, showed significant ($p < .01$) effects attributable to the anxiety variable, but Bartlett's test revealed significant heterogeneity of variance. Since no transformation could be found to ameliorate this situation, the discrimination task conditions variable was collapsed. This source of variation entered into no significant relationships with the dependent variable and is perhaps of scant theoretical interest. Bartlett's test on the resulting four variances indicated no significant heterogeneity of variance ($p > .10$). The 2 × 2 analysis of variance carried out on the collapsed data[2] yielded significant effects associated with anxiety ($F = 8.04$, $df = 1, 43$, $p < .01$) but nonsignificant effects due to order and to the interaction ($p > .20$). Individual $t$ tests of the overall means of the two groups indicate that the HA *S*s show a significant overestimation of the failure picture relative to the success picture ($t = 2.84$, $df = 22$, $p < .01$), but the relative underestimation trend for the LA *S*s is not significant ($t = 1.28$, $df = 24$, $p > .05$).

Since the *S*s in this experiment had MAS scores from essentially the whole range, it was possible to ascertain the magnitude of the relationship between MAS and success-failure scores for size estimation. For the S-First condition, the Pearson $r$ between these two variables was $-.47$ ($p < .05$); for the F-First order, it was $-.25$ (NS). The pooled correlation (obtained by averaging Fisher's $z$ scores) was $-.38$.[3] This significant ($p < .01$) but low correlation is slightly higher than those reported between MAS scores and eyelid conditioning (Hilgard, Jones, & Kaplan, 1951; Spence & Taylor, 1953). Visual inspection of the scatter plots revealed a slight curvalinearity, so correlation ratios (eta) were computed. These turned out to be .67 and .54 for the two conditions. A test for linearity of regression (Walker & Lev, 1953, p. 278) revealed no significant curvalinearity, indicating that the $r$'s reported above are probably fairly good estimates of the true relationship.

TABLE 1

SIZE ESTIMATION MEANS OF THE REINFORCED AND NONREINFORCED PICTURES BY HIGH AND LOW ANXIOUS SUBJECTS

| Anxiety Group | Reinforced Pictures | | Nonreinforced Pictures | |
|---|---|---|---|---|
| | Success | Failure | Building | Trees |
| High | −.181 | −.082 | −.173 | −.155 |
| Low | −.099 | −.149 | −.081 | −.104 |

In order to test for overall differences in size estimation between the two anxiety groups and for differences between the reinforced and control pictures, the data in Table 1 were subjected to an analysis of variance. The only significant source of variation was the interaction involving pictures and anxiety groups ($F = 2.89$, $df = 3,128$, $p < .05$). Tests of the significance of the differences between the means of individual pictures for each MAS group reveal that there is a reliable difference only between the success and failure pictures for the HA *S*s ($t = 2.18$, $df = 24$, $p < .05$), although the difference between Failure and Bl pictures for this group approaches significance ($p < .10$). A separate analysis of variance for the control pictures alone shows no significant differences between groups.

The results for the Fe and Sc pictures were analyzed separately by $t$ tests because the order of presentation of these pictures was not counterbalanced. These show that the size of

[2] Two cells had $N = 12$, one $N = 11$, and one $N = 13$. The mean of the cell was added to the smallest cell and the score closest to the mean of the largest cell was subtracted from that cell. One $df$ was subtracted from total and from the error term in the analysis of variance.

[3] The correlation between these two variables computed for both conditions pooled was $-.31$. This likely underestimates the population value because of the difference between the two conditions in mean success-failure scores.

TABLE 2
SUCCESS MINUS FAILURE SIZE ESTIMATION SCORES AS A FUNCTION OF PSYCHOLOGICAL AND SOMATIC ANXIETY

| Somatic Items | Psychological Items: High | | Psychological Items: Low | |
|---|---|---|---|---|
| | *N* | *M* | *N* | *M* |
| High | 11 | −.032[a] | 8 | .098 |
| Low | 15 | −.162 | 14 | .071 |

[a] A negative score indicates the failure picture was overestimated relative to the success picture; a positive score indicates the reverse.

the Sc picture was not judged differently from any of the others for either group of *S*s, nor was the group difference significant for this picture. The only significant difference involving the Fe picture was its relative overestimation in comparison to the success picture for the HA *S*s ($p < .01$). In view of the multiple comparisons involved and the lack of counterbalancing of order, this difference must be considered only suggestive.

*Somatic and Psychological MAS Items*

Examination of the MAS reveals at least two classes of items: those relating to physiological or somatic manifestations of anxiety and those relating more to purely "felt" or psychological anxiety. In an attempt to increase the precision of the instrument in predicting size estimation performance, the *S*s were scored separately on each of these two kinds of items. The 15 somatic items referred to gastrointestinal symptoms, headaches, blushing, sweating, etc.,[4] while the 35 psychological items referred to nervousness, fears, worries, self-esteem, etc. The *S*s were divided into two groups on each of these subscales. For the psychological items, the highs were those with a score of 10 or greater; for the somatic items, the highs were those with a score of 4 or greater. Table 2 shows the size estimation success-failure difference scores and *N*s according to this twofold classification. A test for independence on the distribution of frequencies yielded a chi square of less than 1, suggesting a lack of relationship between the two sets of items. An analysis of variance was carried out on the size estimation scores, using a method for unequal cell frequencies (Walker & Lev, 1953). The *F* for the psychological items was 13.00, significant below the .001 level ($df = 1, 44$). The *F* for somatic items was 2.43, $.20 < p < .10$, but the difference is in the opposite direction from that expected, the somatic highs showing a relative overestimation of the success picture and the somatic lows showing the opposite trend. The interaction was not significant ($F = 1.05$). Correlations between the psychological anxiety score and success-failure difference were −.62 ($p < .005$) for the S-First order and −.35 ($p < .10$) for F-First, and the pooled correlation was −.50 ($p < .001$).

This analysis indicates that the discriminative power of the MAS with respect to the size estimation task of the present experiment lies in the items referring to psychological anxiety and not in the somatic items.

[4] These items are Nos. 4, 5, 11, 18, 24, 25, 27, 33, 36, 37, 38, 41, 43, 44, and 48 in Taylor (1953, p. 286). Revised forms of items were used (p. 288).

## DISCUSSION

The results of this experiment reveal a difference between high and low manifest anxiety groups in their size estimations of pictures which had been associated with success and failure. This difference was due primarily to a relative overestimation of the failure picture by HA *S*s. Assuming that having failed in a task involving a certain stimulus may increase the "personal relevance" or "importance" of that stimulus for *S*, the data strengthen the generalization that such relevance is an important variable influencing the estimated size of the stimulus. The fact that only the HA *S*s evidenced a significant difference in estimated size between the two pictures indicates that the affective meaning of the failure picture may have been stronger for this group.

The data are also consistent with the suggestion made previously (Zahn, 1959) that a higher anxiety level may have been responsible for the relative overestimation of a failure picture by the good premorbid schizophrenic patients. It is interesting to note, however, in view of the similarities of method in the two experiments, that these patients show a much greater difference between the success and failure pictures than do the present HA

normals. The difference on the same pictures for the patient group is −.272 for eight trials of size estimation and −.367 for the first four trials as compared to −.099 for the HA normals for four trials. This greater difference for the patients would seem to indicate a greater degree of anxiety or perceptual lability or both. In addition, the failure to find any overestimation of the Sc picture in the HA normals—despite their relative overestimation of the failure picture—emphasizes the overestimation of this picture by the good premorbid patients. This would seem to indicate either that normals are not sensitive to symbolic representations of traumatic events as compared to schizophrenic patients, or that the two groups in fact differ in their maternal relationships.

A subsidiary result of this study suggests there may be at least two independent classes of items in the Taylor Scale—here called "somatic" and "psychological" items—which may be measuring different aspects of anxiety. Although in the present case the separation of these items was rather arbitrary and based only on their face validity, the results suggest that the psychological items may carry most of the weight in predicting the effects of failure on performance. Certainly, they suggest a closer look at the individual items of the scale, possibly with factor or cluster analytic methods.

In view of other studies which have found a relationship between positive value and positive reinforcement and size estimation, the question of why the success picture was not also overestimated seems pertinent. One can only speculate in an ad hoc fashion that the operations performed here did not lead to an increase in the value or affective meaning of this picture. Since no special attempt was made to make the discrimination task an ego involving one, it is unlikely that the easy success achieved on the success picture led to any great feelings of accomplishment or of having overcome a challenge, and hence the picture remained quite neutral.

Insofar as the size estimation technique used in the present experiment is comparable to that of previous studies relating size estimation and value, it may be argued that the phenomenon demonstrated is essentially similar. However, a serious question may be raised about the comparability of the present procedure to that traditionally used in the coin studies and others. It can be argued, rightly, that with the relatively long delay between presentations of the standard and variable pictures (4 seconds) in the present procedure, what is being measured is memory and not perception. In fact, if only perception were altered by the variables used here, no differences should be manifest because the adjusted image should also be perceived as larger to the same extent that the standard was.

Granting this to be true, it is still possible to argue that results obtained by the present method are comparable to the previous ones. Certainly, they are comparable to those of Dukes and Bevan (1952) and of Gilchrist and Nesberg (1952), both of which involved a short interval between judgments of a standard and a variable stimulus. However, even in studies in which the standard is physically present while *S* is adjusting the variable stimulus, it would seem probable, perhaps even necessary, that *S* would regard the two stimuli with separate fixations and that the variable would be fixated after the standard. Thus, there is necessarily a finite time interval, albeit a very small one, between fixation of the standard and variable. Of course, more than one comparison is permissible in this procedure, but this should not change the situation appreciably because in any case the variable should be fixated last. Thus, there is a likelihood that the present procedure and those others involving a time interval between presentations of standard and variable stimuli differ from the "traditional" procedure, where both are simultaneously present, mainly in a quantitative fashion: in the time interval between fixating the two. The present procedure may offer less optimal judgmental conditions, but does not differ in any essential way or measure a different psychological function from the traditional procedure.

This analysis advances the suggestion that in all previous work on this problem, it is possible that changes in the memory rather than the perception of the valued or personally relevant objects were being assessed. Perhaps a procedure that would allow for a less equivocal interpretation would successively present the standard stimulus as a neutral figure

(i.e., square or circle) and the variable the value figure. If the value or personal relevance were "built-in" in a manner similar to the one used in the present study, structural properties of the stimuli could be controlled by proper counterbalancing. This method would seem inappropriate for stimuli with "inherent" value, such as coins, hammers and sickles, and mothers, however, unless some other kind of structural factors were utilized to control for differences between standard and variable stimuli.

## SUMMARY

An experiment was performed to test the hypothesis that the size of a picture associated with failure would be overestimated relative to that of a picture associated with success and control pictures primarily by *Ss* with high scores on the Taylor Manifest Anxiety Scale. After practice on a size estimation task, *Ss* were given a discrimination task involving two pictures, responses to one of which were generally followed by "right" and responses to the other by "wrong." The sizes of these pictures and others were then judged by the method of average error.

The results showed that, in accordance with the hypothesis, high MAS *Ss* significantly overestimated the size of the failure picture relative to that of the success picture, but the low MAS group showed an insignificant trend in the opposite direction. In addition, it was shown that the "psychological" anxiety items of the Taylor scale related significantly to the success-failure difference in size estimation, but the "somatic" items did not.

The results are a further confirmation of the generalization that personal relevance is a determinant of size estimation and imply that associating a stimulus with failure makes it more personally relevant for high anxious than for low anxious *Ss*.

## REFERENCES

BRUNER, J. S., & GOODMAN, CECILE C. Value and need as organizing factors in perception *J. abnorm. soc. Psychol.*, 1947, **42**, 33–44.

BRUNER, J. S., & POSTMAN, L. Symbolic value as an organizing factor in perception. *J. soc. Psychol.*, 1948, **27**, 203–208.

DUKES, W. F., & BEVAN, W., JR. Size estimation and monetary value: A correlation. *J. Psychol.*, 1952, **34**, 43–53.

GILCHRIST, J. C., & NESBERG, L. S. Need and perceptual change in need related objects. *J. exp. Psychol.*, 1952, **44**, 369–377.

HARRIS, J. G., JR. Size estimation of pictures as a function of thematic content for schizophrenic and normal subjects. *J. Pers.*, 1957, **25**, 651–671.

HILGARD, E. R., JONES, L. V., & KAPLAN, S. J. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, **42**, 94–100.

LAMBERT, W. W., SOLOMON, R. L., & WATSON, P. D. Reinforcement and extinction as factors in size estimation. *J. exp. Psychol.*, 1949, **39**, 637–641.

SPENCE, K. W., & TAYLOR, JANET A. The relation of conditioned response strength to anxiety in normal, neurotic, and psychotic subjects. *J. exp. Psychol.*, 1953, **45**, 265–272.

TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

WALKER, HELEN M., & LEV, J. *Statistical inference.* New York: Holt, 1953.

ZAHN, T. P. Acquired and symbolic affective value as determinants of size estimation in schizophrenic and normal subjects. *J. abnorm. soc. Psychol.*, 1959, **58**, 39–47.

(Received October 1, 1959)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 463-468

# THE EFFECT OF AROUSED MOTIVATION ON ATTITUDES[1]

HELEN PEAK
*University of Michigan*

EARLIER experiments have demonstrated that attitudes toward certain situations, policies, or actions are predictable from information about the consequences believed to follow from or to be connected with those situations. Thus attitudes toward removing racial segregation were predictable from a combination of information about the perceived consequences (or instrumental properties) of desegregation and judgments of the value of those consequences (Rosenberg, 1956). Furthermore, attitudes toward this issue were changed by altering the instrumental relations (Carlson, 1956; Peak, Morrison, Spivak, & Zinnes, 1956). But if attitudes are thus dependent on the existence of relations between the object of the attitude (viz., Negro segregation) and positively and negatively valued consequences, then attitudes should change if something alters evaluation of those consequences. Suppose, for example, that a student's negative attitude toward his roommate derives in part from a belief that his roommate interferes with studying and thus with making high grades. This attitude might be altered either by changing the belief that the roommate is responsible for the study difficulties (change in perceived instrumentality) or by changing the evaluation of high grades.

The present experiment is an attempt to determine whether attitudes can be changed by intensifying consequences. A number of academic situations were chosen as the attitude "objects" and their perceived relation to making good grades was determined. Then an effort was made to arouse the goal of making good grades by the device of testing immediately after a quiz had been given (arousal condition), on the assumption that the motive of achieving grades is activated more strongly on such an occasion than on an ordinary class day (neutral condition).

I have sketched elsewhere a behavior theory which assumes that the course of response on any occasion depends on a combination of the pattern of *input* at particular points in a psychological structure and on the characteristics of that *structure* (Peak, 1958). The theory also makes the familiar distinction between restructuring or learning and performance (Peak, 1959). Thus change in attitudinal responses may be produced by a modification in the system of relations (structure) on which the response depends, as when the relation of an attitude object to valued goals is altered (Carlson, 1956). Or activating conditions may be changed. The change in attitudinal response predicted in the present study is probably best thought of as the result of a change in input patterns or activation, i.e., in the degree of arousal of the need to make good grades. Whether any such activation results in restructuring can be determined only by testing reactions under activation conditions that were operating at the time that the original measures were made.

## HYPOTHESES

The first hypothesis states that attitudes expressed toward any issue or object ($X$) are a function of the instrumental relations of $X$ to valued consequences.

The second hypothesis is this. Expressed attitudes toward an issue ($X$) are a function of the intensity of arousal of the consequences of $X$, other things being equal. More specifically, attitudes toward any $X$, seen as positively instrumental to attaining good grades, tend to become more positive when there is an increased arousal of the desire to make good grades; with negative instrumentality, attitudes toward $X$ should become less positive under arousal conditions, when positive and negative instrumentality groups are compared to each other and to those judging $X$ to have no instrumental value.

[1] The analysis of these data was completed under contract between the Office of Naval Research and the University of Michigan, Project Nonr 1224(10) NR 171-039. The article was written while the author was in residence at the Center for Advanced Study in the Behavioral Sciences. We are grateful for the skillful assistance with computations by the late Mary Girschick of the center staff.

## METHOD

Various experimental measures were given to two groups of University of Michigan undergraduates during regular sessions of their psychology classes. In Group I, measures were made under arousal conditions during the first session and under neutral conditions in a second session 10 days to 2 weeks later. In Group II, measurements were first made under the neutral condition and later under arousal conditions. These two groups of subjects have been combined in the analysis. A total of 115 women and 85 men took the tests.

The following measures were given in the order indicated.

*Arousal and Neutral Procedures.* In the arousal condition the period began with a 15-minute psychology quiz that had been previously announced for this period. Subjects were stopped promptly after 15 minutes and were told to pick up materials that had been placed at hand ready for use. The quiz was followed by a need Achievement Test and attitude measures. An additional 10-minute session was devoted to the instrumentality sorting procedure in a quiz section usually on the following day.

For Group I the second session was the "neutral condition." The quiz papers had been given back a week or 10 days earlier. The session started with a need Achievement Test, using a set of pictures equivalent to the first set. Then the attitude statements were reacted to and the instrumentality sorting procedure was carried out. In Group II measures were made under this neutral condition during the first session.

*Instructions.* Subjects were told that a full explanation of the overall purposes of the experiment would be made after the study was completed and this was done. The specific instructions for each test were given immediately before the administration of the test. Subjects used codes on all their test forms and no other identification was asked for.

*need Achievement Measures.*[2] Standard instructions for the need Achievement Test were given (McClelland, Atkinson, Clark, & Lowell, 1953, pp. 98–99). Following the projection of each picture on a screen, subjects wrote answers to the four questions used in this standard procedure. It was intended to compare achievement scores under neutral and aroused conditions in the expectation that this measure would reflect an increased arousal of the goal of making good grades.

*Attitude Measures.* Following the achievement measures subjects were asked to respond to a series of statements by indicating degrees of agreement or disagreement with each statement (+4, +3, +2, +1, −1, −2, −3, −4).

One introductory statement (A) was included for practice. Three were designed to explore the subjects' feelings of personal adequacy and three were taken from the California F Scale (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950). The following 11 statements were used to test our Hypotheses.

A. Closing the Library all day Sunday should be protested vigorously by all students and faculty.

1. Students who take books off Library reserves without signing up for them should be called before the College Judiciary and penalized severely.

2. Giving unannounced quizzes is justified and should be general practice.

4. Textbooks should have a practical point of view rather than deal with theory.

5. It is necessary that the discussion method be used in all classes and that less time be devoted to lectures than to discussion.

6. The University should increase the budget so that no class has more than twenty students in it.

8. Every college teacher should be trained to concern himself with the special needs of the individual student rather than spend so much time with the content of the subject taught.

9. Examinations should be abolished because they are rather poor measures of what the student knows.

10. It is desirable that a full time academic load be considered to be 12 credit hours of course work rather than 15 credit hours per semester.

11. It is the responsibility of every student to spend at least 8 hours of study per week on each three hour course.

12. Students should be required to watch their physical condition.

15. Studying at night is better than studying in the daytime.

*Instrumentality Sorting.* Subjects spent 10 to 15 minutes sorting 12 cards on each of which was written a phrase identifying a policy or procedure corresponding to one of the attitude statements that had been responded to on the previous day: for example, "giving unannounced quizzes" (Item 2), "8 hours of study per week for each 3 hour course" (Item 11), "using a textbook with a practical point of view and a minimum of theory" (Item 4). They were instructed to consider the probability that the given procedure or policy would lead to, interfere with, or have no effect on making good grades. Subjects read all the cards before beginning to sort and then placed the cards in three piles, one including those statements which had nothing to do with making good grades, another those that would interfere, and finally those that would help. After this the subjects picked the statements that would help most and put the identifying numbers of those statements in a column headed +5, those next most helpful in Column +4, and so on. Then they placed the most interfering procedures in Column −5, somewhat less interfering in −4, etc.

In arousal sessions, where 15 minutes were occupied with the quiz, the instrumentality sort was administered at the beginning of the next regularly scheduled class discussion period, usually the next day. Before the sorting started, an effort was made to reinstate the arousal conditions of the preceding day by referring to the quiz questions with very general comments in-

[2] We wish to thank J. W. Atkinson for help in the use of the need Achievement procedures and in the scoring of the protocols. We used Pictures A, B, and G in that order for the neutral condition and C, E, and D for aroused. These pictures are regarded as equivalent forms (McClelland, Atkinson, Clark, & Lowell, 1953, Ch. VII).

tended not to answer the questions but to start the subject thinking about his answers.

In the neutral condition need Achievement, attitude measures, and instrumentality sorts were performed in the same session.

*Face Sheet Information.* All subjects filled out an information blank at the end of the arousal session, indicating sex, grade point averages, and information about the last date on which they had taken a quiz and the next day on which they expected a quiz. Subjects who had taken quizzes in some other class on the same day or the day before were eliminated from the sample, as were those who had a quiz scheduled for the following day.

## RESULTS

The results will be reported in four sections. The first considers the relation of instrumentality scores and attitudes (Hypothesis I). The second discusses the relation of change in need Achievement scores to experimental conditions (neutral and aroused) and to attitude change scores with instrumentality held constant. The third section also presents data relevant to Hypothesis II, change in attitude as a function of the arousal of related goals. A final section discusses the relation of instrumentality scores obtained under aroused and neutral conditions.

### *Instrumentality and Attitude—Hypothesis I*

The results of this analysis appear in the first column of Table 1. Judgments of instrumentality under arousal conditions were significantly related in the predicted direction to attitude scores under arousal conditions in 8 of the 11 items for men and 10 of 11 for women. In the case of Item 1 only the men showed a significant relation but the trend for women was in the same direction. For Items 6, 9, and 11 the women showed significant results while the men did not, although the relations for men were in the predicted direction in all but one case.

In short, for men and women all relations but one were in the predicted direction. Where a policy or procedure was seen as facilitating good grades, the attitude toward the procedure tended to be positive; when interfering with making good grades, attitudes tended to be negative. Hypothesis I was strongly supported.

### *need Achievement Correlates*

Interest in the relation between achievement measures and attitude change arose from the expectation that the achievement measure might reflect the level of the need to achieve good grades. It was hoped that achievement scores would increase as a result of changing the experimental conditions from neutral to aroused, and that attitude change in subjects who showed evidence of change in level of n Ach could be analyzed separately from those who did not.

Contrary to expectations, there were no significant differences in need Achievement scores of subjects under aroused and neutral conditions whether men and women were analyzed separately or together. This means, then, that either the need Achievement Test did not measure change in the motivational state aroused by having taken an examination, or there was no change in motivation produced in the group by these conditions.

An analysis of change in attitude scores as a function of change in achievement scores (neutral minus aroused) with instrumentality held constant also produced completely negative results.

Attitude change under aroused and neutral conditions must be examined, therefore. If there is change as a function of conditions but not as a function of achievement scores, it will suggest that the need Achievement Test was not measuring intensity of arousal.

### *Attitude Change as a Function of Instrumentality and Goal Arousal—Hypothesis II*

Since change in need Achievement scores showed no relation to attitude change, Hypothesis II was tested by observing the direction of change in attitude as a function of experimental conditions, given certain judgments of the instrumental properties of the attitude object in leading to or blocking attainment of good grades.

The relation of instrumentality to attitude change (arousal minus neutral scores) is found in Column 2 of Table 1 for men and women separately and combined (C). These probabilities are derived from chi square analyses of instrumentality (positive, zero, and negative) plotted against attitude change (aroused minus neutral). In some instances where number of cases in one of the instrumentality conditions was too small, only two instrumentality categories were used (nega-

TABLE 1

ATTITUDES AND ATTITUDE CHANGE AS A FUNCTION OF INSTRUMENTALITY AND GOAL AROUSAL

| Item | Sex | $p$ (two-tailed) of Chance Relation | | | | | | Instr. (Aroused) and Instr. (Neutral) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Instr. (Aroused) and Attitude (Aroused) (1) | | | Instr. (Aroused) and Attitude Change (2) | | | phi | $p$ (3) | |
| 1. | M | <.01 | (59)[a] | Pre[b] | >.10 | (59) | Pre | .35 | <.01 | (63) |
| | W | >.10 | (101) | Pre | >.10 | (100) | Pre | .49 | <.001 | (107) |
| | C | | | | <.02 | (159) | Pre | | | |
| 2. | M | <.001 | (65) | Pre | >.10 | (65) | Cur[c] | .45 | <.001 | (63) |
| | W | <.001 | (105) | Pre | >.10 | (105) | Cur | .50 | <.001 | (108) |
| | C | | | | >.10 | (170) | Cur | | | |
| 4. | M | <.001 | (59) | Pre | >.10 | (57) | Pre | .35 | <.10 | (63) |
| | W | <.001 | (100) | Pre | >.10 | (100) | Pre | .51 | <.001 | (105) |
| | C | | | | >.10 | (157) | Pre | | | |
| 5. | M | <.001 | (58) | Pre | <.001 | (63) | Pre | .52 | <.001 | (64) |
| | W | <.001 | (106) | Pre | >.10 | (101) | Pre | .50 | <.001 | (109) |
| | C | | | | <.001 | (164) | Pre | | | |
| 6. | M | >.10 | (59) | Pre | <.02 | (59) | Pre | .52 | <.001 | (64) |
| | W | <.001 | (102) | Pre | <.05 | (102) | Pre | .44 | <.001 | (108) |
| | C | | | | <.01 | (161) | Pre | | | |
| 8. | M | <.001 | (62) | Pre | <.05 | (62) | Pre | .44 | <.001 | (61) |
| | W | <.001 | (106) | Pre | <.01 | (106) | Pre | .43 | <.001 | (104) |
| | C | | | | <.05 | (168) | Pre | | | |
| 9. | M | >.10 | (65) | Pre | >.10 | (65) | Pre | .32 | <.02 | (64) |
| | W | <.001 | (107) | Pre | >.10 | (107) | Pre | .62 | <.001 | (109) |
| | C | | | | <.10 | (172) | Pre | | | |
| 10. | M | <.05 | (61) | Pre | >.10 | (61) | Pre | .41 | <.01 | (61) |
| | W | <.001 | (103) | Pre | <.10 | (103) | Pre | .57 | <.001 | (107) |
| | C | | | | >.10 | (164) | Pre | | | |
| 11. | M | >.10 | (64) | | >.10 | (64) | Pre | .43 | <.001 | (62) |
| | W | <.001 | (107) | Pre | <.10 | (107) | Pre | .25 | <.02 | (105) |
| | C | | | | >.10 | (171) | Pre | | | |
| 12. | M | <.01 | (63) | Pre | <.05 | (63) | Pre | .35 | <.01 | (62) |
| | W | <.001 | (107) | Pre | <.05 | (107) | Pre | .60 | <.001 | (107) |
| | C | | | | <.01 | (170) | Pre | | | |
| 15. | M | <.01 | (63) | Pre | >.10 | (63) | Cur | .46 | <.001 | (62) |
| | W | <.001 | (102) | Pre | >.10 | (102) | Cur | .40 | <.001 | (104) |
| | C | | | | >.10 | (165) | Cur | | | |

[a] Number of cases in parentheses.
[b] Pre indicates the relation was in the predicted direction.
[c] Curvilinear relation indicated by Cur.

tive and zero plus positive, or positive and zero plus negative). Results for 9 of the 11 items were in the predicted direction for men; the same number for women and the same number for men and women combined ($p <$ .033, one-tailed sign test). For men 4 of the 9 relations in the predicted direction were significant; none were significant which did not show predicted trends. For women 5 of the 9 were significant when one-tailed tests were used. When results of men and women were combined, 6 of the 9 items showed significant relations in the predicted direction using one-tailed tests. These findings suggest then, that goal-arousal conditions did tend to be associated with increased frequency of positive attitudes toward policies seen as positively instrumental to the goal and with decreased

frequency of positive attitudes toward those issues negatively instrumental.

It may be concluded, therefore, that Hypothesis II was supported unambiguously for six of the items for the sample as a whole. Of the remaining five, three were in the predicted direction. The two which did not show predicted attitude change as a function of instrumentality judgments and of goal arousal were Items 2 and 15. The first of these stated that: "Giving unannounced quizzes is justified and should be general practice." The other was: "Studying at night is better than studying in the daytime." The reason for the deviant behavior of these particular items is not obvious.

*Instrumentality under Aroused and Neutral Conditions*

In the results reported, measurements of instrumentality made under arousal conditions have been employed. It is necessary, therefore, to report the relation between the two sets of instrumentality measurements.

Phi coefficients were determined from contingency tables for each item for men and for women and are presented in Column 3 of Table 1. It is clear that the correlations were all positive and significant. Probabilities are those of the contingency tables from which phi was derived.

Evidence of a possible shift in the direction of the instrumentality judgments was sought in the proportion of cases in which there was positive, negative, and no change in the instrumentality direction. Cases of no change were divided equally between the positive and negative categories and the departure of these frequencies from chance was determined.

In the case of Item 1 (taking books out of Library without permission) the instrumentality ratings became more negative under arousal conditions in men ($p < .10$). Ratings of Item 6 (no class should have more than 20 students) in women became more negative following arousal ($p < .05$); men changed in the same direction but not significantly. Women also saw Item 11 (8 hours of study) as less positively instrumental under arousal ($p < .05$). Men changed in the same direction but not significantly. Finally, Item 12 (maintaining good physical conditions) was judged less positively instrumental ($p < .05$ in men; $p > .10$ in women). The fact that all significant changes under arousal were in the direction of seeing the procedures as less positively effective toward the goal may mean that judgments made immediately after taking the quiz were more "realistic" and that some of the policies referred to in the items did not appear to be as helpful as had been anticipated. As a matter of fact, the correlations between attitude and instrumentality and between attitude change and instrumentality were somewhat higher when instrumentality judgments made under arousal conditions were used rather than those in the neutral condition.

*F Scores*

The three F Scale items and the items concerning the subjects' feelings of adequacy did not have any relation with the variables of this experiment.

## SUMMARY

In each of two sessions 200 undergraduates judged the degree of agreement or disagreement with 11 academic policies. They then judged each of the policies (viz., having small discussion groups instead of lectures) in terms of the degree to which the policy in question either facilitated or interfered with the attainment of good grades; this was known as instrumentality sorting. At the beginning of each of these sessions, subjects reacted to Thematic Apperception Pictures which have been standardized as measures of need Achievement.

In one half of the total sample, the first session followed immediately after a 15-minute quiz which had been announced in advance. This was the arousal condition. For this group the second session 2 weeks or 10 days later was a neutral condition. In the other half of the sample, the neutral condition was established in the first session and the arousal condition in the second.

Hypothesis I stated that the attitude toward a policy or object is a function of the judged consequences of the policy. The hypothesis was strongly supported as it had been in earlier studies.

Hypothesis II stated that the attitude expressed is a function of the degree of arousal of consequences to which the attitude object has an instrumental relation. There were two approaches to the test of this hypothesis. It had been expected that scores on the need Achievement Test would reflect the effect of arousal conditions and that those who increased their achievement scores would show the predicted relation between instrumentality and attitude change more clearly than those whose achievement scores decreased or did not change as a result of arousal. The analyses produced no significant relations between change in achievement scores and experimental conditions or attitude change.

The other test of the hypothesis involved the observation of change in attitude toward a policy from neutral to arousal conditions as a function of judgments of the positive, zero, or negative instrumentality of the policy for making good grades. When tested by this method Hypothesis II was supported. The direction of attitude change was that predicted in a significant proportion of the items: i.e., when a policy was judged to be positively instrumental, attitudes toward it became more positive under arousal than when instrumentality was zero or negative. On the other hand attitudes tended to become more negative as a result of arousal when the instrumentality was negative.

This effect of arousal on expressed attitude is regarded as demonstrating the role of so-called performance or activation factors in determining attitude responses.

## REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.

CARLSON, E. R. Attitude change through modification of attitude structure. *J. abnorm. soc. Psychol.*, 1956, **52**, 256–261.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.

PEAK, H. Psychological structure and psychological activity. *Psychol. Rev.*, 1958, **65**, 325–347.

PEAK, H. Attitude change and attitude structure. Final Report, 1959, University of Michigan, Contract Nonr 1224(10) NR 171–039.

PEAK, H., MORRISON, H. W., SPIVAK, M., & ZINNES, J. L. Factors in resistance to attitude change, personality factors. Technical Report No. 1, 1956, University of Michigan, Contract Nonr 1224(10) NR 171–039.

ROSENBERG, M. J. Cognitive structure and attitudinal affect. *J. abnorm. soc. Psychol.*, 1956, **53**, 367–372.

(Received October 2, 1959)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 469–473

# AUDIENCE STRESS, PERSONALITY, AND SPEECH[1]

HARRY LEVIN, ALFRED L. BALDWIN, MARY GALLWEY,[2] AND ALLAN PAIVIO[3]
*Cornell University*

SPEECH appears to be an especially sensitive indicator of a person's feelings. Even excluding the content of what a person says, we often attribute a variety of traits or momentary states to him on the basis of such paralinguistic qualities as length of utterance, pauses, inflection, intonation, stress, etc. For example, the statement that a person is verbose connotes a host of personal characteristics, including that he welcomes a lengthy exposure to his listeners. Also, one of the most ubiquitous inferences from voice qualities is that the speaker is tense or nervous or ill at ease. Brown (1958) calls the bases of such inference the "expressive symptoms" of language.

The present research deals with two aspects of children's speech: the amount of time the child spends in talking and the number of speech errors he makes during his discourse. Our purpose is to predict each of these speech behaviors from situational factors—the number of people listening to the child talk—and from two personality dispositions which we have labeled "Exhibitionism" and "Self-Consciousness" (hereafter referred to as Ex and S-C).

Speaking in public may be thought of, generally, as a stressful situation. If people are forced into this condition and given the option of shortening or prolonging their public exposure, we may reasonably expect them to terminate their performance more rapidly than if the condition were less public. On the other hand, there are differences among people that influence how easily they tolerate public exposure and, in fact, to what extent they actually seek publicness. These individual variations complicate the simple situational prediction. The measures of Ex and S-C were designed to ascertain more precisely the ways in which the public exposure might be handled by various subjects.

The two dispositions may be thought of as approach and avoidance tendencies, respectively, to public performance. The Ex items deal quite specifically with public performance or the exposure of one's products to public appraisal. The briefer S-C scale taps a more general disposition of apprehensiveness about one's evaluation by others. The latter scale correlates substantially with such general anxiety scales as the Children's Manifest Anxiety Scale and Sarason's Test Anxiety Scale. The development of the two scales has been described in detail elsewhere (Paivio, Baldwin, & Berger, in press). They have been validated against volunteering for performance in skits, and their child rearing antecedents have been studied (Paivio, 1959).

If children are given the option of prolonging or terminating their public performance, highly exhibitionistic children for whom this occasion is pleasureable should extend their discourses compared to their less exhibitionistic peers. Conversely, self-conscious subjects, for whom talking to an audience is discomforting might be expected to exercise their option to endure such exposure as briefly as they can.

This prediction suggests that the two variables are negatively correlated. While they do have opposite effects on willingness to submit to public exposure, they are not the opposite ends of a bipolar scale.

Based on a number of samples of children, exhibitionism and self-consciousness have been found to be only minimally or not at all correlated. A child, thus, may wish to perform in public, yet be apprehensive about this exposure. This phenomenon is reminiscent of stage fright among actors. We reasoned, consequently, that subjects who score high on both scales would be experiencing an approach-avoidance conflict during their public performance—they want to get out of the situation and yet are drawn toward it. Such conflict may be contrasted with the less complicated apprehension felt by children who

[1] This study was supported by grants from the National Institute of Mental Health (M-901). We wish to thank Irene Stein, Katrin Norris, and Janet Von Hendy for their help with this study.

[2] Now at Washington State University.

[3] Now at the University of New Brunswick.

are very self-conscious but not exhibitionistic. A mode of resolution of the latter state is readily available to the subject: he can make his performance brief and thus leave the stressful situation. Such a straightforward termination of conflict is not available to the high-high group, because the act of leaving increases the positive valence of remaining before the public, which, in turn, calls out their self-consciousness, etc.

Besides influencing the length of their public performance, the conflict is also presumed to result in more speech errors. Since the subject who exhibits a high degree of approach and avoidance tendencies is under an unresolved stress situation in public, we expect his voice to exhibit more symptomatic signs of stress than the voices of subjects who have a different pattern of the two personality variables.

In summary, this study tests the hypothesis that on the average the public performance of speech is a stressful situation to children in our society and as such results in both briefer discourses and more expressive symptoms of uneasiness in the voice. The extent, and possibly the direction, of the influence of the public situation on a particular person, however, should depend on the degree to which he possesses two personality characteristics, Ex and S-C. Children high in Ex and low in S-C should find public speaking less anxiety provoking and more attractive than those with the opposite pattern. The subject who possesses a high degree of both of these dispositions should, however, be under greatest conflict during public performance, a state that should be reflected in a high rate of speech errors.

## METHOD

### *Subjects*

The subjects were 48 children, equal numbers of boys and girls, aged 10 to 12 years. They were all from a single elementary school in a lower middle class neighborhood.

### *Procedure*

*Measures of Exhibitionism and Self-Consciousness.* Details about the nature of the items in the two scales and their factor analytic structure are given in detail in another publication (Paivio et al., in press). Ex consists of 13 items, each of which contributes a weight of +1 to the scale score. Examples of items that have high factor loadings on the scale are: "I like to tell a story in front of the class"; "I like to show things I make to other children."

The S-C scale has six items. Two examples of these items are: "I often wonder what others think of me"; "Other people hurt my feelings easily."

Test-retest reliabilities on these scales are .62 for Ex, and .45 for S-C, based on a sample of 209 children comparable in age and schooling to the subjects in the present study.

The items were embedded in a total of 79 and administered to 115 children in four classrooms. The final sample of 48 subjects was selected to represent four groups of boys and four of girls according to whether they scored above or below the medians for the total sample tested on the Ex and S-C scales. That is, one group was high Ex, high S-C, another high Ex, low S-C, etc.

*The Public Speaking Situation.* Each subject told stories under both "private" and "public" conditions. Before his first formal story the subject was trained in story telling in the following way. The experimenter took him from his classroom to another room where he was given a sentence and told to make up a story about it. During this preliminary session, the experimenter answered questions, asked the usual projective queries, and prompted the subject if he paused for so long that it seemed he would not go on without help. In order to help the child get the idea of a full story, the experimenter pointed out after the story was finished, that he had told a story which included a beginning, what happened, and how it came out.

Private and public experiences were counterbalanced so that following the preparatory session the child told two stories to the same experimenter or went to another room where he told two stories to an audience of six adults; the original experimenter was not present during this condition. The adults in the audience were instructed to listen interestedly and to respond naturally, but not to interact directly with the children. The two sessions for any single child were about one week apart. All stories were tape recorded.

Four beginnings used to instigate the stories were randomly assigned across children and conditions and were worded to match the sex of the child:[4]

1. A boy gets up in front of the class to make a speech.
2. A boy is a good actor. There is going to be a school play.
3. The teacher asks a boy to come to the front of the room and work at the blackboard.
4. A boy has prepared a long report on social studies. The teacher asks him to come to the front of the room and lead a discussion about the report with the children.

Two coders working independently, scored the following categories from the tapes:

1. Length of speaking (seconds)

[4] The story instigators, which are all obviously concerned with public performance, were developed in an earlier study to yield roughly comparable contents. The analysis of the content of the oral stories in the present study showed no consistent variation with either the personality or situational variables.

2. Number of words uttered
3. Number of stutters
4. Prolongations (like "sp-e-e-ech," or "my-y-y")
5. Shifts (starting a word or sentence and shifting before the utterance is finished, as "the b-girls want" or "the girls went—want")
6. Sighs
7. Pauses (any cessation of speech for 2 seconds or more was scored and the actual length of the pause was timed)
8. Vocal segregates (these are the ubiquitous "uhs" around words)
9. Catches in the voice

The lowest correlation between scorers on any one category is .80 and the highest is above .90.

In summary, this study involves a $2^5$ factorial design made up of the following classifications: Ex, S-C, audience condition, sequence, and sex. The dependent variables are expressive speech characteristics symptomatic of stress, and length of voluntary exposure in story telling.

## RESULTS

We are concerned with the prediction of two dependent variables—speech errors and length of speech—from personality and situational variables. The two dependent variables are uncorrelated in our data.

### *Length of Talk*

Based on a previous study (Levin & Baldwin, 1958) and the pretests of this study we hypothesized that highly exhibitionistic subjects should speak for a longer period of time, given the opportunity, than their less exhibitionistic peers. Further, if we continue our assumption that the audience condition is more stressful than the more private one, and we recall that the child controls, to a great extent, the length of his exposure, we may predict briefer literary creations in the audience than in the private condition. This last anticipated finding may be qualified by interactions with the personality measures in that highly exhibitionistic subjects should be least discomfited by the public condition and highly self-conscious children, most.

The mean seconds speaking according to the various predictor categories are given in Table 1. From the analysis of variance the relevant findings are: Public ($F = 25.60$, $p < .001$); Public × Ex ($F = 4.78$, $p < .05$); Public × S-C ($F = 5.19$, $p < .05$).[6] Almost

[6] Several other significant interactions that were not predicted but do not influence the interpretation of the above two findings are not being considered.

TABLE 1
MEAN NUMBER OF SECONDS OF SPEAKING
(N = 12 in each cell)

| Exhibitionism | Self-Consciousness | Condition | |
|---|---|---|---|
| | | Public | Private |
| High | High | 208 | 275 |
| | Low | 227 | 265 |
| Low | High | 141 | 341 |
| | Low | 168 | 231 |

every one of the 48 children told briefer stories to the six listeners than to the single experimenter. Overlaid on this general decrease, however, is the finding that highly exhibitionistic children decreased their stories *least*. Subjects who are very self-conscious, on the other hand, showed the greatest shortening of their stories in the public compared to the single auditor condition. It will be recalled that the selection of the research groups obviates any correlation between the two personality measures. The two interactions with "public" summate so that the most spectacular decrease in time of exposure to the listeners occurred for the low Ex–high S-C group: a decrease of over 3 minutes, on the average.

Although we had no definite predictions about the highly self-conscious children who were not exhibitionistic, their long discourses in the private condition invite speculation. It may be that these usually reticent children overrespond to the warmth and individual attention of the private condition as it was designed in this study.

### *Speech Errors*

For the two major personality variables—Ex and S-C—we predicted that the subjects who score high on both scales would be under greatest stress in a public situation and that the stress would produce a high incidence of speech mistakes. Further, the subjects who are highly Ex but not S-C should be least upset by public performance and therefore make few speech errors. Also, we anticipated a situational main effect indicating greater stress when the child was talking in public than when making up stories to the experimenter alone.

The means relevant to testing these hypotheses are presented in Table 2.

TABLE 2
MEAN RATES OF SPEECH ERRORS
(N = 12 in each cell)

| Exhibitionism | Self-Consciousness | Condition | |
|---|---|---|---|
| | | Public | Private |
| High | High | .071 | .055 |
| | Low | .032 | .041 |
| Low | High | .045 | .057 |
| | Low | .063 | .056 |

The frequency of any one of the speech categories was usually so low that we calculated an overall speech error score by summing all of the categories and dividing the sum by the total number of words uttered, so that the score we are dealing with is "rate of speech errors." The distribution of this score was normalized by an arc sine $\sqrt{X}$ transformation (the means in Table 2 are not transformed).

An analysis of variance of these data indicated the following results: Ex × S-C ($F = 5.12$, $p < .05$); Ex × S-C × Public ($F = 5.14$, $p < .05$). The basic hypothesis is confirmed. Most speech errors are made by children who are conflicted over public performance, i.e., high in both Ex and S-C, and fewest errors are made by exhibitionistic children who show little apprehension about exposure (the high Ex and low S-C group). In addition, whether the child is performing before many or one observer puts a sort of astigmatic glass over this finding. In private, the strong implication is that only the high-low group is different from the others—they err least. In public, the conflicted (high-high) group is far higher in mistakes than the others.

When we look farther than the predicted error rate in the high-high group, the array of means in Table 2 raises some perplexing questions. The second largest error rate was made by subjects who are low on both characteristics. Our impression is that these children are generally unresponsive and apathetic. For them, speech errors—perhaps especially pauses—are not so much a response to situational stress as a generally slow behavioral pace. The rather close correspondence between the speech errors in both the public and private conditions favors our speculations. It suggests also, that future studies of paralinguistic behavior include a "normal" baseline derived from various natural speech situations.

In general, our hypotheses about speech errors and length of exposure are confirmed, with the exception that Ex did not appear as a significant main effect with seconds talking. It is interesting to note that when "number of words" rather than seconds is the dependent variable, highly exhibitionistic children utter more words irrespective of the audience condition.

## DISCUSSION

Our theoretical assumptions at the outset of the study were that we might isolate two tendencies toward public performance. We thought of "exhibitionism" as an approach tendency toward such exposed situations and "self-consciousness" as an avoidance tendency. The tendencies are, in our theory, dispositions which are elicited when the person contemplates or is actually in a public situation.

These assumptions tend to be confirmed by the speech errors which we take as symptoms of stress. People who possess both these tendencies to a strong degree should be most conflicted (stressful) under an audience situation and this is actually the case. Also, the exacerbation of the larger audience magnifies the effect.

However, the results for length of exposure raise the question of whether there is ever an unadulterated seeking and preference for public performance. There is, for example, much folklore about experienced performers who repeatedly suffer agonies of stage fright. Similarly, even the highly exhibitionistic subjects told briefer stories in the public as compared to the private situation, although it is true that their reduction in time was considerably less than other children. The point, nevertheless, is that they did suffer less exposure when reciting to a group. Perhaps, the Ex variable might better be thought of as relative weights of approach and apprehension over public behavior.

Ex and S-C are conceptually related to a person's self-esteem. A person may score high on Ex because he sees the probability of success as high and just the converse for the self-conscious person. Yet for the exhibitionistic person a special kind of success may draw him to audience situations and that is success or

achievement *in public*. This reasoning has lead us to make a distinction in conventional achievement motivation theory.

Many scorable instances of achievement according to McClelland, Atkinson, Clark, and Lowell (1953) involve achievement in public. In fact public approval for goal attainment appears to be a dominant motive for some people. It seems to us useful therefore, to make a distinction between pure achievement, where public performance is not relevant, and exhibitionistic achievement, where reaching the goal is simply instrumental to public approbation. We speculate that highly exhibitionistic children are such instrumental achievers. One clue to this interpretation is that they deliver longer stories, in number of words, in both audience conditions than do subjects who score at the lower end of this scale. The present findings set us to try to explicate the distinction between pure and exhibitionistic achievers in our further research.

## SUMMARY

This study was designed to test the effects of a situational factor—the presence or absence of an audience—and two personality variables—exhibitionism and self-consciousness—on the length of speech and on the number of errors in speaking. Forty-eight children, aged 10 to 12 years, told two stories to the experimenter alone and two to an audience of six adults. The story beginnings were given to the subjects. The findings were:

1. Almost every subject told a longer story in the private compared to the audience situation.

2. Highly exhibitionistic children were least influenced in this respect whereas self-conscious children reduced the length of their stories drastically before the six observers.

3. Children who were both highly exhibitionistic and self-conscious made the most speech errors.

4. This finding was magnified under the audience condition.

## REFERENCES

BROWN, R. *Words and things*. Glencoe, Ill.: Free Press, 1958.

LEVIN, H., & BALDWIN, A. L. The choice to exhibit. *Child Develpm.*, 1958, **29**, 373–380.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.

PAIVIO, A. Child rearing antecedents of audience sensitivity. Unpublished doctoral thesis, McGill University, 1959.

PAIVIO, A., BALDWIN, A. L., & BERGER, S. Measurement of children's sensitivity to audiences. *Child Develpm.*, in press.

(Received October 7, 1959)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 474-479

## *CASE REPORTS*

# THE CONTEXTUAL ASSOCIATES OF SUN AND GOD IN SCHREBER'S AUTOBIOGRAPHY[1]

JULIUS LAFFAL

*Veterans Administration Hospital, West Haven, Connecticut*

DANIEL PAUL SCHREBER was a doctor of jurisprudence and a prominent judge at the end of the last century, who, after becoming psychotic, published an account (1903) of his delusions in the belief that he was making a contribution of value to science and the knowledge of religious truths. Without having known Schreber personally, Freud (1948) undertook on the basis of the published autobiography, to trace the etiology of Schreber's delusions, and in his study formulated the now famous paradigm of homosexuality and projection in paranoia. Interest in the Schreber case has continued high among psychoanalysts (Baumeyer, 1956) and a number of papers have been published on various aspects of the case. One notable study is that of Macalpine and Hunter (1955) who also published an English translation of Schreber's autobiography (1955). These authors differed with Freud's interpretation of the importance of homosexuality in the paranoid system, and maintained that a wish for ambisexuality and self-contained procreative powers was at the root of Schreber's illness. They brought the difference between their view and Freud's to a focus in contesting Freud's account of the symbolism of the sun in Schreber's delusions (Macalpine & Hunter, 1955):

> . . . if the sun was not a father symbol for Schreber, Freud's theoretical deductions only partially cover the facts. The passive homosexual wish-fantasy, ultimately derived from his infantile relation to his father and evidenced in his relation to the sun, could not afford the whole explanation, but would throw light only on one aspect, however important, of his psychosis and the content of his delusions.
>
> All the evidence goes to show that the sun, far from representing only the father, mirrored Schreber's own ambisexuality, being both male and female (p. 378).

The issue of the symbolism of some word or referent would seem one that could be dealt with only by speculation and argument. However, an approach to the matter is possible, which uses internal consistency of the writer's (or speaker's) language as a measure of the similarity of themes. Thus, if Schreber used the same terms in writing of the sun as he did in writing of father figures, this could be taken as evidence that his associations to the two were alike and that in effect they meant the same thing to him. If on the other hand, Schreber wrote about the sun the way he wrote about women, this could be taken as evidence for the Macalpine and Hunter view.

In the present study an effort is made to ascertain the extent of the similarity between the way Schreber wrote about sun and certain other words. The theoretical rationale of the technique to be described has been outlined elsewhere (Laffal, unpublished) and is not requisite in detail to follow the discussion. Briefly, the technique is a development from the viewpoint that meaning may be described as a matrix of associations: verbal associations where one is concerned primarily with verbal meaning and that similarity of meaning may be described in terms of similarity or overlap of associations. This view forms the basis of several studies by Bousfield, Whitmarsh, and Danick (1958). In free speech the contexts of a given word are taken to be the associated matrices of the word, that is, the sets of words which would tend to appear in varying frequencies whenever that word occurred in speech. To make such contextual analyses feasible, it is necessary to categorize the words in the contexts on the basis of synonymity or similarity of reference. The basic criterion for using a category would be the simple one that it be capable of subsuming a fair number of the subject's (or the language's) words.

Many words may be categorized only if one arbitrarily focuses on some special feature of the response. The pertinence of a given word to one category rather than another may be established by definition; however, an analysis of similarity of contexts like the one outlined above would be the ultimate validation criterion to determine if the verbal response belonged in a given class. That is, to the extent that the contextual associates of word $x$ overlapped the contextual associates of other words in Class A, then to that extent, word $x$ would belong in Class A.

[1] This report is part of a larger study of Language Distortions in Schizophrenia. The study is supported by United States Public Health Service Grant M-2020.

This paper has been reviewed in the Veterans Administration and is published with the approval of the Chief Medical Director. The statements and conclusions are the result of the author's own study and do not necessarily reflect opinion or policy of the Veterans Administration.

TABLE 1

CATEGORIES OF WORDS IN THE CONTEXTUAL ANALYSIS OF SCHREBER'S LANGUAGE

| | | |
|---|---|---|
| *agree-*disagree | *forward-*back | *possess-*want |
| animal-nature | free-confine | *real-unreal |
| *begin-*end | good-bad | reason-absurd |
| *big-little | *help-*hurt | sacred-profane |
| *body feature-*body function | *hollow-hill | *see-*hear |
| calm-*upset | hot-cold | *separate-*join |
| clean-*dirty | in-out | sex-homosex |
| *come-go | *individual-*group | *sharp-blurred |
| commodity-money | law-crime | *some-all |
| cure-*illness | lead-follow | speak-*write |
| *durable-*change | *living-dead | *strong-weak |
| *earth-sea | male-female | *structure-*ruin |
| easy-difficult | *near-far | true-false |
| eat-drink | new-ancient | *up-*down |
| *essential-trivial | open-shut | young-mature |
| fast-*slow | *play-*work | |

Note.—All of the categories in the list were initially applied in the scoring of Schreber's language. However, only the 42 categories identified by asterisks were used in correlating contexts with each other.

A set of 94 categories, grouped roughly into opposition pairs, which had been developed from the study of the speech of a schizophrenic patient, were applied to the analysis of Schreber's autobiography.[2] Heading words for these categories are listed in Table 1, with asterisks showing those categories which were ultimately used in the analysis of the Schreber material. The heading words merely suggest the nature of the words in categories and are not adequately descriptive of the categories.

Unfortunately the third chapter of Schreber's autobiography, which discussed intimately the members of his family, was deleted by censorship, so that there are practically no references to Schreber's father in the autobiography. This makes a direct comparison of speech about father with speech about sun impossible. However, there are some indirect ways of approaching the problem. It is clear in Freud's discussion, that God is the father figure par excellence. Macalpine and Hunter (1955) do not accept this view and consistently with their position about the sun, say: "... his [Freud's] conviction that God and the sun were 'nothing but' father symbols for Schreber; he ignores their obvious female significance..." (p. 378). Some fairly definite predictions arise from these differences of view.

If Freud's contention that sun and God are both father symbols is correct, we ought to find a strong relationship in the way Schreber writes about both of these, and both of them ought to be more strongly related to the way Schreber writes about male figures than the way he writes about female figures.

If the Macalpine and Hunter view is correct, then we ought also to find a strong relationship between the way Schreber writes about sun and God, but both of these ought to be strongly related to the way Schreber writes of females. Their view would seem to entail the prediction that the sun-female and God-female relationship would be at least as strong as, and possibly stronger than the corresponding male relationships.

Two papers by Niederland (1951, 1959) adduce considerable evidence based on a study of Schreber's father's life and works, that the accusations which Schreber directed against God, especially relating to body torture, were also appropriate to Schreber *père*. The older Schreber wrote extensively on the theme that in the earliest years of the child's life a maximum of pressure and coercion should be used in the inculcation of physical and emotional restraint. He devised various mechanical instruments for enforcing correct body posture, not only in waking but during sleep, and he believed in strict disciplinary training measures, including corporal punishment. Niederland (1951) maintains that not only God, but Schreber's physician, Flechsig, who is mentioned often in the autobiography, was a father symbol for Schreber: "First Flechsig, and then God, is reinstated in the place of father" (p. 591). Freud does not make this direct relationship between Flechsig and God and Flechsig and father. Although Freud (1948) does say that "in the patient's mind 'Flechsig' and 'God' belonged to the same class" (p. 49), Freud ultimately identified

[2] The Macalpine and Hunter translation (1955) was the actual text used in the analysis. It was felt that the fine problems of meaning which usually arise in translations would not seriously affect the analysis applied here. This, however, remains unexplored.

TABLE 2

PEARSON PRODUCT-MOMENT CORRELATIONS BETWEEN CONTEXTS OF KEY WORDS

$N = 42$ (categories)

| | | male | | female | | God | | Flechsig | |
|---|---|---|---|---|---|---|---|---|---|
| | | Set 1 | Set 2 | Set 1 | Set 2 | Set 1 | Set 2 | Set 1 | Set 2 |
| male | Set 1 | | | | | | | | |
| | Set 2 | .706[a] | | | | | | | |
| female | Set 1 | .667 | .561 | | | | | | |
| | Set 2 | .457 | .437 | .566[a] | | | | | |
| God | Set 1 | .226 | .475 | .166 | .420 | | | | |
| | Set 2 | .451 | .534 | .413 | .352 | .679[a] | | | |
| Flechsig | Set 1 | .679 | .645 | .576 | .504 | .446 | .598 | | |
| | Set 2 | .556 | .662 | .516 | .319 | .552 | .636 | .737[a] | |
| sun | | .523 | .330 | .678 | .356 | .136 | .444 | .511 | .397 |

[a] Reliabilities.

Schreber's homosexual feelings toward Flechsig as derivatives of those toward an older brother:

> The feminine phantasy, which aroused such violent opposition in the patient, thus has its root in a longing, intensified to an erotic pitch, for his father and brother. This feeling, so far as it referred to his brother, passed, by a process of transference, on to his doctor, Flechsig; and when it was carried back on to his father a settlement of the conflict was reached (p. 50).

These remarks about Flechsig suggest that the contexts surrounding the physician's name might also provide some data about the male-father symbolization of God and sun. Freud maintains at least that Flechsig and God were both the subject of homosexual fantasies of the patient. In this sense both stood in the relation to him of a dominant male figure. Niederland goes further than Freud in equating Flechsig with father. If the analyses of Freud and Niederland are correct, we ought to find a strong relationship in the way Schreber writes about Flechsig and God, and both of these ought to be more strongly related to the way Schreber writes about male than the way he writes about female. Macalpine and Hunter are not specific in interpreting the symbolism of Flechsig in Schreber's psychology, but they appear to suggest also some element of identity with God (Macalpine & Hunter, 1955):

> Endowed with this divine power as head of the 'Nerve-Institute,' Flechsig was supernatural and partly a soul [§ vii–xii, pp. 56–57] in direct contact with God [§ 82, p. 111]. Herein lies the easy passage from Flechsig to God [§ 49, p. 95] and 'God Flechsig' [§ 82] (p. 396).

## PROCEDURE

All passages in the text of Schreber's autobiography, which contained reference to only one of the words sun, God, male, female, or Flechsig,[3] were marked for transcription. A passage was defined as a sequence beginning three lines before a line containing a key word and ending three lines after a line containing the key word. A roughly equivalent number of such passages was taken out for each key word. It happened that the word sun occurred alone least often. The total number of passages available for sun was used to determine the number of passages selected for each of the other key words since it was desired to have approximately equal numbers of scored responses in the ultimate comparisons of contexts. For all key words other than sun (for which no additional passages were available), an additional equivalent number of passages was taken out for purposes of reliability study.

The words in the passages were categorized in accordance with the categories in Table 1.[4] An illustration of the categorizations for a passage relating to God is:

> big strong join body function begin
> Only the increased power of attraction of nerves caused by
> strong help speak nature, unreal
> it enabled and facilitated communication with supernatural
> strong cure difficult upset
> powers. It seems psychologically impossible that I suffer only
> unreal back all unreal
> from hallucinations. After all, the hallucination of being in
> speak lead, sacred dead, go unreal reason
> communication with God or departed souls can logically only
> in group group group illness
> develop in people who bring with them into their morbidly

[3] In reviewing the passages which had been used in the analysis, it was discovered that the word Flechsig appeared in a few passages of all of the other words. However, no effort was made to repeat the analysis without these passages.

[4] Examples of the kinds of words scored under each of the categories, and of scored passages from Schreber's autobiography, have been deposited with the American Documentation Institute. Order Document No. 6546 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

upset body function strong agree, sacred
excited nervous state an already secure faith in
lead, sacred living, dead unreal
God and the immortality of the soul. This, however, was not
individual speak begin write
so in my case, as mentioned at the beginning of this chapter.
speak unreal unreal reason true
Even so-called spiritualist mediums may be considered genuine
unreal unreal down, weak
seers of spirits of the inferior kind.

In order to study reliability of the frequencies of associated categories, for those words in which enough passages were available, the lines were randomly divided into two sets, each containing an equivalent number of lines in which the key word appeared, and an equivalent number of lines one, two and three lines above or below the key word line. There were thus two sets each of distributions of responses associated with male, female, God, and Flechsig, and one set of responses associated with sun.

By inspection of the distributions of categories it was determined that there were some categories which rarely occurred in Schreber's writing, and some categories which occurred consistently and with great frequency. Such categories were eliminated from the analysis as contributing only to a general increase of all correlations and not to the discrimination between Schreber's verbalizations on the key words. Further, the specific categories under which man, woman, sun, God, and Flechsig were scored were also eliminated from the comparisons, since selection of the critical passages had been made on the presence of one and the absence of the other categories. The total number of categories was in this way reduced to 42, which are identified by asterisks in Table 1. The number of associated words remaining for each key word ranged from 192 to 258.

## RESULTS

In Table 2 are shown the reliabilities of each of the associated contexts for all the key words, except sun, for which no reliability was available, and the intercorrelations of all of the sets of contexts with each other. The correlations are Pearson product-moment correlations obtained by comparing frequencies of occurrence of the 42 categories identified by asterisks in Table 1. Table 3 shows the average correlation between the sets of contexts for all key words. These averages were arrived at by taking all the intercorrelations, two for sun with the other words, and four for all the other words with each other, and dividing by the number of correlations.

The contexts for female are the least reliable ($r = .566$). The reliability of the male contexts is .706, for the God contexts it is .679, and for the Flechsig contexts it is .737.

It will be noted that for the contexts of male, female, God, and Flechsig, sets of four intercorrelations are available for each pair. The test of

TABLE 3
AVERAGE CORRELATIONS RANKED FROM HIGHEST TO LOWEST

| Correlation between: | Correlation |
|---|---|
| male-Flechsig | .636 |
| God-Flechsig | .558 |
| male-female | .531 |
| female-sun | .517 |
| female-Flechsig | .479 |
| sun-Flechsig | .454 |
| male-sun | .427 |
| male-God | .422 |
| female-God | .338 |
| sun-God | .290 |

significance takes advantage of this fact. Thus, to test the hypothesis that the way Schreber talks about God is more like the way he talks about male than the way he talks about female, the 8 correlations involved (4 for God-male and 4 for God-female) may be compared with each other, making 16 comparisons. The God-male correlations are greater in 13 of the comparisons, the God-female correlations are greater in 3 of the comparisons. By chance each set of correlations would have been expected to be greater than the other 8 times. A sign-test analysis (Johnson, 1949) of this relation shows that the God-male correlations are significantly more frequently greater than the God-female correlations, with a probability less than .05.

With respect to significance tests of the relationships between sun and male and sun and female, a problem arises. Since only one sun context is available, only two sun-male and sun-female correlations are possible. However, as a test of the different viewpoints of Freud, and of Macalpine and Hunter, comparisons may be made, for the Freud prediction, between sun-male and sun-God on one side, and sun-female and God-female on the other; for the Macalpine and Hunter prediction, between sun-female and sun-God on one side, and sun-male and God-male on the other.

In neither instance is the difference significant. Inspection of the correlations in Table 2 shows that the contexts of sun are not related in any consistent way either with the contexts of God and male, or God and female. The average correlation for sun and female is higher than for sun and male, and of the four comparisons available between the correlations, three are in favor of the sun-female correlation, one in favor of the sun-male correlation. Not enough data are available, however, for a significance test.

Certain other tests of significance may be done, bearing on the relationship between the way Schreber writes of Flechsig and the way he writes

about God, male, and female. The relationship between Flechsig and male is significantly greater at the .01 level than the relationship between Flechsig and female, Flechsig and God, and God and male. In addition, the relationship between God and Flechsig is significantly greater at the .05 level than the relationship between God and male.

## DISCUSSION

The findings are rather clear with respect to some of the major hypotheses tested. Several predictions based on Freud's position are supported. Thus, the contexts of God in Schreber's autobiography are significantly more like the contexts of male than of female; the contexts of Flechsig and male are significantly more alike than those of Flechsig and female. The relationship between the contexts of God, male, and Flechsig also suggest that in the slight difference of opinion between Niederland and Freud, the Freudian position may be stronger. Niederland saw both Flechsig and God as father symbols in Schreber's writing. Freud maintained that while both Flechsig and God were in the same class, God was the father substitute, and Flechsig stood in the relationship of a brother. It will be recalled that there are not available in the autobiography enough contexts of father to make direct comparisons possible. In this analysis we have had to settle with contexts of male to test the various hypotheses. Schreber writes about Flechsig more like he writes about male than about God. He writes about God more like he writes about Flechsig than about male. If the matrix of associations around Flechsig is seen as composed of male and father words with male words predominating, and the matrix of associations around God as composed of father and male words, with father words predominating, this could account for the strong relationship between Flechsig and male, the moderately strong relationship between Flechsig and God, and the comparatively weak relationship between God and male.

With respect to the relationship between sun and God as symbols, the findings support neither Freud's nor Macalpine and Hunter's position. Freud saw sun and God in the autobiography as father symbols, and Macalpine and Hunter ascribed feminine qualities to both of them. Unfortunately sun is not written about enough in the autobiography for tests of the reliability of the distribution of its contextual associates, and for enough comparisons with other contexts to provide meaningful tests of significance.

The comparisons made in the present study combine two predictions from both Freud and Macalpine and Hunter: that sun is like God and that both are either like male or like female. These comparisons do not achieve statistical significance, and inspection of the correlations of sun with God suggests that one reason is that sun is not very highly correlated with God.

Table 3 shows female and sun to be more highly related, on the average, than male and sun. Of the four direct comparisons between these two correlations in Table 2, three are in favor of the female-sun correlation, and one in favor of the male-sun correlation. The comparison is statistically inconclusive, but the tendency of the data is in the direction of the prediction derived from Macalpine and Hunter's view that sun is a feminine symbol.

Since the word sun is used relatively infrequently in the autobiography as compared to words like God, Flechsig, male, and female, it is likely that both Freud and Macalpine and Hunter were forced to speculate more freely on its symbolism and relationship to other aspects of Schreber's thinking. In each case some striking context of the word seems to have been taken as the key to its interpretation. Thus, Macalpine and Hunter refer to the fact that Schreber calls the sun a whore. In addition they pointed out that the sun is of feminine gender in German and that there is folk-mythological support for the idea of the sun as an ambisexual thing. Freud laid emphasis on the fact that Schreber himself identified the sun sometimes with the "upper" and sometimes with the "lower" God. Freud also drew on evidence from other cases, from literature, and from mythology to support his assertion that the sun is a father symbol for Schreber. It is possible that Freud and Macalpine and Hunter, in relying heavily on universal interpretations of the sun as a symbol, and having relatively little textual material available in the actual autobiography, failed to evaluate correctly its significance in Schreber's system. It is beyond the scope of this paper to attempt such an analysis, but the findings point to the caution necessary in introducing universal symbolism into the verbalizations of an individual. The complexity of the sun symbolism is well illustrated in a paper by Abraham (1953) who while adhering to the position that the sun is a father symbol, pointed out the many ways in which it is also a female symbol.

The technique that has been used in this study to explore the predictions derived from Freud and Macalpine and Hunter bears some further comment. Much trial and error has gone into the evolution of the method, and it cannot yet be described as a polished procedure. There remain numerous uncertainties about the best way of classifying some words, about whether to broaden or narrow some of the categories, about how many lines of a

text should be considered a passage relevant to the key word under study, and so forth. In actually categorizing words, there are many instances in which one's definitions call for a certain categorization of a word but where the categorization seems odd and strained. The procedure itself will undoubtedly undergo numerous transformations as experience accumulates, and such changes may make possible more refined and comprehensive analyses of the same texts. The purpose of the procedure is to provide an objective way of comparing segments of free-flowing language within the theoretical framework that associative similarity is an index of meaning similarity. In the present paper the method has been applied to examining a dispute about the symbolism of certain words in Schreber's autobiography. However, the method is quite general, and it should be applicable wherever problems of similarity of written or spoken stretches of speech arise.

## SUMMARY

An objective procedure of contextual analysis was applied to test opposing predictions derived from Freud's (1948) and from Macalpine and Hunter's (1955) studies of the Schreber case. Freud maintained that both God and sun were father symbols for Schreber; Macalpine and Hunter contended that God and sun were ambisexual symbols and stressed their feminine connotations. The contextual analysis involved categorizing words appearing in close association with the key words God, sun, and Flechsig and in close connection with male and female words, and correlating the distributions of categorized responses with each other. It was found that the way Schreber wrote about God was significantly more like the way he wrote about male than the way he wrote about female. Schreber's way of writing about Flechsig was also significantly more like the way he wrote about male than about female. However, it was found that the way Schreber wrote about sun was not similar to the way he wrote about God. Schreber wrote about sun more the way he wrote about female than about male, but this distinction was statistically inconclusive.

## REFERENCES

ABRAHAM, K. Restrictions and transformations of scoptophilia in psychoneurotics. (Originally published 1913) In, *Selected papers of Karl Abraham*. New York: Basic Books, 1953. Ch. 9.

BAUMEYER, F. The Schreber case. *Int. J. Psycho-Anal.*, 1956, **37**, 61–74.

BOUSFIELD, W. A., WHITMARSH, G. A., & DANICK, J. J. Partial response identities in verbal generalization. *Psychol. Rep.*, 1958, **4**, 703–713.

FREUD, S. Psychoanalytic notes upon an autobiographical account of a case of paranoia. (Originally published 1911) In, *The complete psychological works of Sigmund Freud*. Vol. XII. London: Hogarth, 1948.

JOHNSON, P. O. *Statistical methods in research*. New York: Prentice-Hall, 1949.

MACALPINE, IDA, & HUNTER, R. A. Discussion of the Schreber case. In D. P. Schreber, *Memoirs of my nervous illness*. (Trans. by Ida Macalpine & R. A. Hunter) Cambridge: Bentley, 1955.

NIEDERLAND, W. G. Three notes on the Schreber case. *Psychoanal. Quart.*, 1951, **20**, 579–591.

NIEDERLAND, W. G. Schreber: Father and son. *Psychoanal. Quart.*, 1959, **27**, 151–169.

SCHREBER, D. P. *Denkwürdigkeiten eines Nervenkranken*. Leipzig: Mutze, 1903.

SCHREBER, D. P. *Memoirs of my nervous illness*. (Trans. by Ida Macalpine & R. A. Hunter) Cambridge: Bentley, 1955.

(Received November 3, 1959)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 480-481

# A CLINICAL NOTE ON THE SPECIFIC NATURE OF AN "ORAL" TRAUMA[1]

BERTRAM P. KARON[2]
*Philadelphia Psychiatric Hospital*

THE apparent discrepancies between clinical case studies and experimental investigations are disturbing to those of us who believe that there is only one psychology of man, whatever the source of the data.

The importance of oral experiences and traumata in the production of character traits and neurotic and psychotic symptoms has been stressed by Freudian psychoanalysis (Abraham, 1954; Fenichel, 1945, pp. 62–66). The role of these determinants seems to be manifest in case history material, and in many recent modifications of Freudian theory, oral problems are given even greater emphasis (Bergler, 1949; Klein, Heiman, & Money-Kyrle, 1955; Silverberg, 1952). Nonetheless, the bulk of the experimental work, as Orlansky (1949) has noted, does not seem to confirm the importance suggested by theory. It may therefore be worthwhile to discuss the relevant aspects of a patient whose adult symptoms showed clear indications of "oral" pathology but seemed to have no childhood basis for such problems. The nature of his oral trauma may help to clarify the disparity between psychoanalytic reconstructions and experimental findings.

The term "orality" is used in psychoanalysis to refer to the mouth and its functions—especially eating, drinking, sucking, and biting—to the dynamic characteristics connected with these (e.g., dependency is seen as being connected with the experience of being fed), and to the first psychosexual "stage" of early infancy, when these functions of the mouth are assumed to be the most important activities in the life of the child. The term is thus used to refer both to problems which originate in this early period of life and to problems having to do with mouth functions. According to Fenichel, one would expect oral pathology to be prominent in the symptoms of someone who had been excessively deprived of oral gratification, excessively indulged in the oral period, or excessively indulged in the oral period and excessively frustrated in the succeeding stage.

Both in theoretical discussions and in attempts to investigate the importance of "oral" experiences, attention has frequently been focused upon whether or not the child was breast-fed or how early he was weaned. More recently, it has been realized that an abrupt weaning at any age is more traumatic than a gradual one, no matter when it takes place (Sears & Wise, 1950).

Destructive oral experiences have been said many times to play an essential role in the development of schizophrenia (Bellak, 1948, 1958; Fenichel, 1945, pp. 415–452; Rosen, 1953; Sechehaye, 1955). A 32-year-old male chronic paranoid schizophrenic patient who illustrates this contention has been described in a previous paper (Karon & Rosberg, 1958), but the basis for his oral trauma was not spelled out in that brief account.

This patient's delusions were such that any psychoanalytically oriented psychotherapist would have concluded that he suffered severe problems of an obviously oral nature and that he must have undergone an oral trauma of considerable impact. For example, the patient was afraid of being poisoned. He would not drink milk at the beginning of the therapy and confided that: "The Athenian girls are laughing at me. They say their breasts are poisoned."

One day a woman with a rather well-endowed bosom walked by him. He made frantic wiping gestures from his side. Afterwards he said: "She approached me with her breast, and I got a pain in my side."

Despite these symptoms, it was possible to verify that this patient had been breast-fed for a full year and had been gradually weaned. Where, then, could there be an oral trauma?

On one occasion, his mother (who had read some of the literature of psychoanalysis) said to him: "Didn't I give you enough milk? Didn't I give you enough to drink?" The patient stared into space and replied: "The cow gave her calf milk and then kicked it. She shouldn't do that. It's something that happened hundreds of times in the history of the world."

In this delusion of the cow who kicked her calf (the patient, of course, denied that it had anything to do with him) is the key to the oral trauma. It did not consist of being deprived of the breast or of being weaned too early or too abruptly. The oral trauma consisted of a sequence of interactions: The child got fed and then got hurt. The "poisoned"

[1] This note was completed during a United States Public Health Fellowship and is based on work previously done.

[2] The writer wishes to express his appreciation to Jack Rosberg, who collaborated in the study of the patient described here, and to John N. Rosen, under whose supervision the patient was treated.

milk seems not to be due to any characteristic of the milk itself, but to the fact that the mother resented any demand upon her to feed the child so that immediately after feeding him, she got angry. The child then felt the impact of her hostility and was hurt. "Poison" is simply something you eat, and you get hurt afterwards.

This sequence—child eats, mother gets angry, child feels hurt—was not something which happened once or twice in the patient's life. It was not something that happened to him only during a specific period of childhood. It was something that started when he was an infant and continued all his life. Even at the age of 32, it was observable. The mother took care of the treatment home[3] for a period of time so as to save expenses. She seemed to be making every conscious effort to help her son. The therapist was startled to notice that after every meal she cooked for her son, she had an argument with him. If she did not cook the meal, there was no argument. Thus, it was not merely a single childhood experience, but the continuous re-enactment of the trauma that prevented it from ever being overcome.

It must not be supposed that this was in any way a conscious process for the mother. Consciously, she was doing everything she could for her son. She was unaware of her constant pattern—feed, get mad—which seemed obvious to an observer. Each evening the content of the argument dealt with something else.

Her emotional inability to feed is exemplified by the fact that after raising two children and after many years of married life, her cooking was barely palatable, not only according to the patient but according to the therapist's judgment as well. Her needs were exemplified in the very beginning of the treatment when she complained of her obviously sick son: "He always took and never gave; it's about time he gave a little." Actually, the patient had been forced to give in a number of different ways—his gifts to his mother, his writings, his paintings, etc. (Karon & Rosberg, 1958).

Since it is typical of the schizophrenogenic mother that she cannot tolerate the demand to feed, to take care of, to nurture, it is likely that this repeated pattern of activity—feed, get mad—is frequently present in such cases. The important point is that the oral trauma is not merely a single childhood event, but the continuous re-enactment of the traumatic sequence.

[3] The patient, while under therapy, lived in a private house with treatment personnel and with his mother.

Without being aware of it, the mother continuously imprinted the notion that after eating comes the pain. Is it any wonder then that the patient is afraid of being poisoned?

Thus, the apparent discrepancy between clinical case studies and experimental findings seems to be due to the fact that "oral trauma" has been interpreted as referring to a single specific event—usually weaning. This is not entirely the fault of the investigators since psychoanalytic theory is usually presented as if the key to symptom and character formation lies in such isolated events. The present case suggests that the important "oral" variables are the persisting patterns of interactions, including the mother's involuntary and unconscious reaction to the demand for nurturance. Isolated events, like weaning, are easily influenced by "enlightened" child care education or by conscious good intentions without removing the destructive character of the total relationship. It seems likely that when attention is directed to the continuing relationship between mother and child, the discrepancies between clinical and experimental data will disappear.

## REFERENCES


ABRAHAM, K. On the first pre-genital stage of the libido. In, *Selected papers on psychoanalysis.* New York: Basic Books, 1954. Pp. 248–279.

BELLAK, L. *Dementia praecox.* New York: Grune & Stratton, 1948.

BELLAK, L. *Schizophrenia.* New York: Logos, 1958.

BERGLER, E. *The basic neurosis.* New York: Grune & Stratton, 1949.

FENICHEL, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.

KARON, B. P., & ROSBERG, J. Study of the mother-child relationship in a case of paranoid schizophrenia. *Amer. J. Psychother.*, 1958, **12**, 522–533.

KLEIN, MELANIE, HEIMAN, PAULA, & MONEY-KYRLE, R. E. *New directions in psychoanalysis.* London: Tavistock, 1955.

ORLANSKY, H. Infant care and personality. *Psychol. Bull.*, 1949, **46**, 1–48.

ROSEN, J. N. *Direct analysis.* New York: Grune & Stratton, 1953.

SEARS, R. R., & WISE, G. W. Relation of cup-feeding in infancy to thumb-sucking and the oral drive. *Amer. J. Orthopsychiat.*, 1950, **20**, 123–160.

SECHEHAYE, MARGUERITE A. *Symbolic realization.* New York: International Univer. Press, 1955.

SILVERBERG, W. V. *Childhood experience and personal destiny.* New York: Springer, 1952.


(Received February 29, 1960)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 482-483

## *CRITIQUE AND NOTES*

# MALE AND FEMALE SUBJECTS IN PERSONALITY RESEARCH

EARL R. CARLSON AND RAE CARLSON
*Michigan State University* *East Lansing, Michigan*

THERE is a sizable accumulation of evidence today that males and females react differently to many social situations and differ in many personality characteristics. Not only are there mean differences on a number of psychological measures, but often a relationship appears for one sex group but not for the other. Potential sex differences are omnipresent, but are largely ignored in even the best of current research on personality and social problems.

The present report is directed to three problems we see in current research: the imbalance in sampling in research, the paucity of tests for sex differences, and the inadequacy of reporting of the sex composition of research samples. An individual investigator, of course, must select his sample to meet specific needs of the problem. Typically, a study is criticized only if the investigator states an obvious overgeneralization. Beyond this, however, it is important at times to view the research literature as a whole to note possible implied overgeneralizations as a function of consistent biases entering into the selection of samples. From such views we can often redirect our efforts, and explicitly check for the limits of generalizability of early findings.

The 298 empirical studies reported in the 14 consecutive issues of the *Journal of Abnormal and Social Psychology*, Vol. 56 (1958), Number 1 through Vol. 60 (1960), Number 2, were selected as a sample of current research on personality and social phenomena. The few studies based upon single cases, nonhuman subjects, or reanalyses of previous research data were not included in this sample. Tabulations were made of the sex composition of the samples of subjects studied, the completeness of reporting of sex composition, and the reports of tests for, and findings of, sex differences.

The results show an overwhelming use of males as compared with female subjects. Of a total of 298 studies reported, 113 (38%) used male subjects exclusively, and 15 studies (5%) used only female subjects. A total of 108 studies reported using both male and female subjects, and for 65 studies the sex of subjects was unknown. The studies using both male and female subjects were fairly well balanced; in 36 studies (33%) an equal number of males and females were used, and in 20 studies (19%) the proportions of males and females were not reported.

Obviously, sex differences could be tested in only those 108 studies using subjects of both sexes. In only 32 studies (30%) were tests of sex differences reported. But in 22 of these cases (69%) sex differences were found at a level of statistical significance.

Several implications follow from these findings. First, males are selected far more often than females as research subjects. In studies of many psychological processes this bias may lead to no problems, but on a number of personality and social measurements sex differences do occur. As readers of research reports we can say to ourselves that of course we realize that only males were used, and that generalizations to females must depend upon further research. It is not only difficult to retain such contingencies in our recollection of studies we have read, but, we venture to guess, readers are largely unaware of sex composition in reported studies—even when this information *is* reported.

How often are sex differences actually tested? If it is correct that all *tests* for differences are reported, then with 69% of these showing significant differences it would follow, with little question, that differences do in fact exist in numerous cases, but were overlooked by the researcher. If many tests were made but never reported, then the reported differences could be expected by chance alone. Either of these alternatives would seem more plausible than the possibility that unreported tests precisely parallel the sex differences reported.

Even for a reader who is cautious in generalizing results, interpretation is ambiguous when the sex composition of the sample is incompletely reported. In 20 of 108 studies reporting both male and female subjects the proportions are unknown. In 62 studies (21% of all studies) a reader is totally unaware of the sex of the subjects, although many of these studies presumably included both men and women. In several studies the $N$ can be discerned only from a thorough reading of the tables. We suspect that all readers will assume that Yale undergraduates and naval recruits are male, as we have done in the tabulations for this paper. There may be other similar situations where the authors

assume, perhaps too optimistically, that readers can infer the sex of subjects from other information given.

Mean sex differences may or may not affect tests of psychological relationships between variables. In cases where test ceilings restrict the range of scores for one group, the variance for this group may be reduced to the point where a relationship cannot be tested. In a substantial number of studies a relationship holds for one sex but not for the other. A more serious problem arises in studies using extreme groups, where mean differences between sexes may lead to an imbalance of sex in the groups selected for intensive analysis. Some studies separate sex distributions and select high and low groups separately from each (e.g., Goldstein, 1959). Other studies do not (e.g., Taylor, 1958) and the resulting findings may stem from other correlates of sex than the variable under study.

Obviously many characteristics of an individual may interact with psychological relationships under study. A researcher cannot hope to do more than control or test for the effects of more than the most significant of these factors. We suggest that for two reasons, at least, sex should be included routinely as an explicit variable in studies where such a comparison is possible. First, sex differences do run rampant through our literature, and no one can deny that they are often important. Secondly, sex is an easy factor to measure, and sex differences are easily tested and reported. There seems little excuse not to include them as a standard procedure.

## REFERENCES

GOLDSTEIN, M. J. The relationship between coping and avoiding behavior and response to fear-arousing propaganda. *J. abnorm. soc. Psychol.*, 1959, **58**, 247–252.

TAYLOR, J. The effects of anxiety level and psychological stress on verbal learning. *J. abnorm. soc. Psychol.*, 1958, **57**, 55–60.

(Received May 2, 1960)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 484-485

# PERSONALITY SIMILARITY, POSITIVE AFFECT, AND INTERPERSONAL ATTRACTION[1]

CARROLL E. IZARD
*Vanderbilt University*

RECENT studies have shown that perceived similarity (Newcomb, 1956), assumed similarity (Fiedler, 1958), and actual similarity (Izard, 1960) are significant correlates of interpersonal attraction. The latter paper postulated that interpersonal positive affect—defined as the expression of favorable feeling, self-involving interest, acceptance or esteem in relation to another person—was a significant factor in much interpersonal behavior and a key determinant of interpersonal attraction. The role of actual personality similarity was assumed to be that of facilitating the mutual expression of positive affect. The paper argued that persons with similar affective characteristics tend to find the same interpersonal situations satisfying and that they realize more accurate interpersonal perception and communication. The finding of the previous study that there was significant personality profile similarity between pairs of mutual friends was considered as support for these arguments.

Like the previous study, the present one assumed that interpersonal positive affect is a key determinant of interpersonal dynamics and that personality similarity facilitates the mutual expression of positive affect. The hypothesis was that subjects and their unilateral sociometric choices have significantly similar personality profiles prior to acquaintance, while subjects and their sociometric rejections do not.

## METHOD AND RESULTS

Edwards' (1954) Personal Preference Schedule (EPPS), the personality measure utilized in assessing profile similarity, was administered to an entire freshman class on entering college. Six months later 47 of the female students in this group were given a sociometric form requiring them to list the three most likeable and the three least likeable girls in their class. The instructions emphasized that each subject should list the three whom she personally felt were most likeable and should avoid listing people on the basis of general popularity. They were requested to exclude from consideration anyone whom they knew prior to entering college.

Twenty-five of the subjects made independent (mutually exclusive) sociometric choices. Twelve of the 25 were randomly selected for pairing with classmates they listed as most likeable and the 13 remaining subjects were paired with the classmates they listed as least likeable. As shown in Table 1, the analysis of variance technique for comparing profiles of pairs (Izard, 1960) indicated that prior-to-acquaintance personality profiles were significantly similar for subjects and their sociometric choices ("most likeable" classmates) but not for subjects and their sociometric rejections ("least likeable" classmates).

TABLE 1

ANALYSIS OF VARIANCE COMPARING PERSONALITY PROFILES OF PAIRS BASED ON SOCIOMETRIC CHOICES OF MOST LIKEABLE AND LEAST LIKEABLE CLASSMATES

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| Between Variables | 14 | 475.75 | |
| Variables × Groups | 14 | 15.98 | |
| Variables × Pairs (Sociometric choices) | 154 | 22.25 | 1.45** |
| Variables × Pairs (Sociometric rejections) | 168 | 16.96 | |
| Variables × Subjects (Sociometric choices) | 168 | 15.38 | |
| Variables × Subjects (Sociometric rejections) | 182 | 25.27 | |

** $p < .01$.

## DISCUSSION

Izard's (1960) earlier study showed that personality profile similarity was a significant correlate of the mutual friend relationship. That study did not show whether the personality similarity existed prior to the friendship or was a result of it.

The present study demonstrated that actual personality similarity was an antecedent of unilateral sociometric choice (nomination as most likeable) but not of sociometric rejection (nomination as least likeable). This finding lends further support to the assumption that personality similarity facilitates the expression of interpersonal positive affect. One aspect of interpersonal positive affect consists of favorable feeling or liking for another individual, precisely the factor that was measured by the sociometric form employed in this study.

[1] This study was supported by Vanderbilt-ONR Contract Nonr 2149(03), C. E. Izard, Principal Investigator. The opinions and conclusions do not necessarily reflect those of the United States Department of the Navy.

REFERENCES

EDWARDS, A. L. *Manual, Edwards Personal Preference Schedule.* New York: Psychological Corp., 1954.

FIEDLER, F. E. *Leader attitudes and group effectiveness.* Urbana: Univer. Illinois Press, 1958.

IZARD, C. E. Personality similarity and friendship. *J. abnorm. soc. Psychol.*, 1960, **61**, 47–51.

NEWCOMB, T. M. The prediction of interpersonal attraction. *Amer. Psychologist*, 1956, **11**, 575–586.

(Early publication received July 6, 1960)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 485-487

# THE STABILITY OF COGNITIVE CONTROLS[1]

RILEY W. GARDNER AND ROBERT I. LONG

*Menninger Foundation*

In a recent monograph, Gardner, Holzman, Klein, Linton, and Spence (1959) developed a conception of dimensions of cognitive control and described a first study of relations between previously defined control principles. However, no adequate studies of the reliability of measures of cognitive controls have yet been published.

A recent study provided an opportunity to retest a group of subjects who had been given criterion tests for several control principles approximately 3 years earlier (mean interval, 36.6 months). The cognitive controls involved (see Gardner et al., 1959) are: Equivalence Range, Scanning, Leveling-Sharpening, and Constricted-Flexible Control.

## METHOD

At the end of 3 years following the initial testing of a group of 80 female subjects, 38 were available for retesting. At the time of the initial testing, these 38 subjects were students at a small midwestern university and ranged in age from 18 to 22, mean age, 19.3.

The control principles involved and the tests and scores used to represent them are summarized below.

*Equivalence Range.* Object Sorting Test: An exact replication was performed. The score is the number of groups formed by the individual in free sorting of 73 objects.

*Scanning.* Size Estimation Test I: The original size estimation score used to represent this control principle was a presumed correlate of the subject's scanning strategy: average error in judging the sizes of plain circular disks. Results of subsequent studies of the Scanning principle (Gardner, 1959, in press) supported the impression of Gardner et al. (1959), however, that constant error is a more appropriate measure than average error to represent the effects of scanning on relative apparent magnitude. An exact replication of judgments of three 40-millimeter disks (gray, 7 gm.; black, 7 gm.; heavy gray, 67 gm.) was performed. In this test, the subject held the disks in her left hand while she adjusted a variable circle of light to be equal to it in size. Scores were constant error in judging each disk and constant error for the entire test.

*Leveling-Sharpening.* Schematizing Test: An exact replication of this test was performed. It should be noted, however, that this test is not repeatable in the same sense as the others discussed here. Effective measurement of assimilative interaction among percepts and memory traces in this test—which involves 150 size judgments of squares ranging from 1.3 to 13.7 inches—requires that the subject be naïve as to the range of sizes and the sequence of presentation. The gradual progression of stimulus sizes (over 35 minutes) is from smaller to larger. A subject who has experienced the full range of sizes approaches retesting with considerable information concerning the general structure of the test. Some subjects have an "aha" experience late in initial testing, when they realize suddenly that they are lagging far behind the trend of size increase. Such subjects abruptly adjust their judgments upward, at times with accompanying written or vocal comments. Other subjects remain unaware of their lagging throughout, even though they may judge the 13.7-inch square as "five inches." Scores used were percentage ranking accuracy, mean percentage increment error (see Gardner et al., 1959), and a recently developed "lag" score, which consists of the slope defined by the logarithms of the actual means of the last nine series of 15 judgments minus the slope of the logarithms of the subject's judgments. This new score omits judgments in the first series and is based equally on the remaining series means. The increment error score is based on the mean percentage increment of the last nine series means over the mean of the first series and hence is overly dependent on the mean of the first series of judgments.

*Constricted-Flexible Control.* Color-Word Test: Exact replication of all three parts of this test (reading color names, color strips, and colors printed in incongruous color names) was performed. Scores consisted of the reading times for each part and the interference score: the difference between the subject's time on Part III and the time predicted on the basis of the regression of time for Part III on time for Part II.

## RESULTS

Means and standard deviations for all tests and measures are presented in Table 1. Means of all constant error scores for Size Estimation Test I are significantly lower on retest ($t < .01$), as

[1] This investigation was supported by Research Grant M-2454, from the National Institute of Mental Health, United States Public Health Service.

TABLE 1
MEANS AND STANDARD DEVIATIONS OF TEST AND RETEST SCORES

| Score | N | Mean | | Standard Deviation | |
|---|---|---|---|---|---|
| | | Test | Retest | Test | Retest |
| Object Sorting Test | | | | | |
| Number of Groups | 34 | 17.1 | 15.7 | 7.8 | 6.8 |
| Size Estimation Test I | | | | | |
| Constant Error, Gray Disk | 33 | 2.85 | 2.02 | 1.75 | 1.43 |
| Constant Error, Black Disk | 33 | 3.15 | 2.15 | 2.22 | 1.50 |
| Constant Error, Heavy Gray Disk | 33 | 2.95 | 2.02 | 2.04 | 1.71 |
| Constant Error, Total Test | 33 | 2.96 | 2.06 | 1.87 | 1.46 |
| Schematizing Test | | | | | |
| Percentage Accuracy | 38 | 66.87 | 64.21 | 14.75 | 13.16 |
| Mean Percentage Increment Error | 38 | 73.39 | 67.31 | 37.19 | 38.89 |
| Lag | 38 | .0158 | .0153 | .0215 | .0192 |
| Color-Word Test | | | | | |
| Reading Time, Color Names | 33 | 41.4 | 41.7 | 6.4 | 6.3 |
| Reading Time, Colors Alone | 33 | 56.6 | 56.4 | 7.1 | 6.8 |
| Reading Time, Color-Word | 33 | 97.4 | 90.5 | 12.2 | 11.5 |

evaluated by $t$ tests for correlated measures. None of the other test-retest means are significantly different.

Pearson correlations between the test and retest scores are presented in Table 2.

Additional findings bearing upon the reliability of these measures include the following:

In the original testing, $r$'s between constant error scores for the three disks used in Size Estimation Test I were: gray:black, .81; gray:heavy gray, .80; black:heavy gray, .82. In the retest, these $r$'s were .90, .80, and .82, respectively.

In the original testing, the $r$ between the accuracy and mean percentage increment error scores for the Schematizing Test was −.22. In the retest, it was −.37. The $r$'s between the accuracy and lag scores were −.26 and −.18 for the two testings.

In the original testing, the $r$ between reading time for the colors-alone and color-word sections of the Color-Word Test (used to obtain the interference score) was .64. In the retest, this $r$ was .61.

TABLE 2
PEARSON CORRELATIONS BETWEEN TEST AND RETEST SCORES

| Score | N | r |
|---|---|---|
| Object Sorting Test | | |
| Number of Groups | 34 | .75*** |
| Size Estimation Test I | | |
| Constant Error, Gray Disk | 33 | .64*** |
| Constant Error, Black Disk | 33 | .57*** |
| Constant Error, Heavy Gray Disk | 33 | .63*** |
| Constant Error, Total Test | 33 | .68*** |
| Schematizing Test | | |
| Percentage Accuracy | 38 | .52*** |
| Mean Percentage Increment Error | 38 | .36* |
| Lag | 38 | .45** |
| Color-Word Test | | |
| Reading Time, Color Names | 33 | .54*** |
| Reading Time, Colors Alone | 33 | .75*** |
| Reading Time, Color-Word | 33 | .75*** |
| Interference | 33 | .55*** |

* $p < .05$.
** $p < .01$.
*** $p < .001$.

## DISCUSSION

The test-retest correlations observed for this group of subjects offer support for the assumption that cognitive controls, as measured by these procedures, are relatively enduring features of cognitive organization. The conception of "low rate of change" processes does not imply exceptionally high reliability, particularly in the present group of relatively young subjects. All the subjects involved in this study underwent major changes of life situation during the interval: 19 of the total $N$ of 38 had been graduated from their university, all but 13 were married, and 4 were pregnant at the time of retesting. All these factors could have contributed to a reduction of $r$'s between test and retest scores.

Except for the significant reduction of constant errors in Size Estimation Test I—which may imply an increase in extensiveness of scanning in the group as a whole (Gardner, 1959; Gardner et al., 1959)—none of the retest means is significantly different from that in the original testing. Apparently, the cognitive controls involved in response to three of the four tests were relatively stabilized at the time of the initial testing.

The lower reliability coefficient for the interference score of the Color-Word Test than for reading times on the colors-alone and color-word

sections of the task is attributable to the fact that reliability of the interference score is a function of the reliabilities of both scores used to compute it, plus the correlation between them.

As anticipated, test-retest correlations are lowest for the Schematizing Test scores. This does not necessarily imply that individual differences in assimilative interaction among percepts and traces are more variable over time than the other cognitive functions sampled. As noted earlier, some subjects, more than others, become aware of the actual range during initial testing. The somewhat higher $r$ for the lag score suggests that it may provide a more effective estimate of assimilative interaction than the earlier mean percentage increment error score.

## REFERENCES

GARDNER, R. W. Cognitive control principles and perceptual behavior. *Bull. Menninger Clin.*, 1959 **23**, 241–248.

GARDNER, R. W. Cognitive controls of attention deployment as determinants of visual illusions. *J. abnorm. soc. Psychol.*, in press.

GARDNER, R. W., HOLZMAN, P. S., KLEIN, G. S., LINTON, HARRIET B., & SPENCE, D. P. Cognitive control: A study of individual consistencies in cognitive behavior. *Psychol. Issues*, 1959, 1, No. 4.

(Early publication received July 18, 1960)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 487–488

# PRESTIGE EFFECTS IN BODY-SWAY SUGGESTIBILITY

J. P. DAS

*Ravenshaw College, Cuttack, India*[1]

THIS study is concerned with the effect of the prestige of the experimenter (*E*) on the body-sway of the subject (*S*). Two orthogonal factors of suggestibility, primary and secondary, have been distinguished by Eysenck (1947), who classifies body-sway suggestibility under primary suggestibility.

## METHOD

Four prestige levels were introduced. Three of these represented three hierarchical levels of prestige, the fourth was of neutral prestige value. The body-sway suggestions of the four *E*s representing the four prestige levels were tape recorded. Their manner of reading out the suggestion was made as similar as possible, the loudness of their suggestions was kept more or less equivalent by the volume adjuster during playback. The four *E*s were: I. Head, Psychology Department, II. Laboratory Assistant, III. Attendant of the Department, and IV. Unidentified. Before each one spoke, the author's voice introduced the speaker giving his name and designation. The unidentified one was introduced as: "Now you are going to hear some one speak; who it is I don't know, for the recording was done when I was absent."

The four prestige levels (I, II, III, and IV) presented in a sequence, admit of 24 permutations. Two *S*s were taken for each permutation, thus requiring 48 *S*s. Static ataxia of *S* for 60 sec. was measured before body-sway suggestions were introduced. Brief intervals not exceeding 60 sec. occurred between the termination of one suggestion and the beginning of the next. Suggestions of any *E* lasted for 2.5 min. During these intervals *S*s sat down in a nearby chair. *S*s were tested individually. The maximum forward sway under such prestige level was recorded.

Forty-eight male college students of the Psychology Department, aged 16–23, were used as *S*s. All *S*s were new to the experiment. All of them knew the Head of the Department, Laboratory Assistant, and the Attendant personally.

To measure body-sway, two sliding pieces of small aluminum plates were mounted on a ruler. Through the sliding plates passed a fine thread. To the part of the thread between the two plates was fixed a tiny piece of metal which could be moved manually, but would not move by striking the metal plates. The thread passed through a pulley at each end of the ruler which was fixed to the wall in a horizontal position. Attached to one end of the thread was a little weight while the other end was to be attached to *S*'s collar by means of a safety pin. When *S* swayed, the tiny metal would strike the piece of aluminum plate either to the left or to the right side according to whether *S* swayed forward or backward. The ruler was 40 in. long, so the aluminum plates were brought to the twentieth in. which served as a central point before sway was recorded. Maximum forward and backward sways could be measured easily by noting the last positions of the left and right side plates immediately after the suggestions were over.

## RESULTS AND DISCUSSION

The means and standard deviations for amount of sway under the suggestions of the four *E*s differing in prestige are shown in Table 1. A fall was scored as 7 in., for beyond that a person would always lose his balance. An analysis of variance was performed taking into account the four prestige and the four order effects. The order effects refer to whether any prestige level was presented

[1] Full address: Postgraduate Department of Psychology, Ravenshaw College, Utkal University, Cuttack 3, India.

TABLE 1
BODY-SWAY (IN INCHES) UNDER DIFFERENT PRESTIGE LEVELS
($N$ = 48)

| | Levels of prestige | | | |
|---|---|---|---|---|
| | I | II | III | IV |
| Mean | 2.21 | 1.90 | 1.75 | 1.66 |
| *SD* | 2.96 | 2.70 | 2.49 | 2.37 |

TABLE 2
ANALYSIS OF VARIANCE FOR BODY-SWAY SCORES UNDER HIGH AND LOW PRESTIGE CONDITIONS

| Source | *SS* | *df* | *F* |
|---|---|---|---|
| Total | 667.716 | 191 | |
| Between | 22.277 | 1 | 6.558* |
| Within | 645.439 | 190 | |

* $p < .05$.

in the first, second, third, or fourth place. For instance, Prestige Level I was presented first to 12 *Ss* and similarly in the second, third, and fourth places to different sets of 12 *Ss*. However, neither the main effects nor the interaction was significant. Since the Laboratory Assistant, Attendant, and Unidentified person together can be regarded as *Es* of low prestige value in comparison with the Head of the Department, it was thought justified to compare Levels II, III, and IV combined with Level I. Accordingly a simple analysis of variance was made (Table 2) neglecting the order effect, and the $i^2$ was significant ($p < .05$).

In order to see the differential effect of prestige on relatively suggestible and nonsuggestible *Ss* their respective scores were further analysed. The first ten high- and low-scoring *Ss* were chosen on the basis of their total suggestibility scores under the four prestige levels. It was predicted that the high suggestibility group would show the differential effect of prestige, whereas the low suggestibility group would not do so. The sign test (Siegel, 1956) was used on the scores of these 20 *Ss*. Comparisons were made only between prestige Levels I and II, I and III, and I and IV, since their differences alone had been previously found to be significant in an earlier sign test. None of these differences were significant for the low-suggestibility group. The difference between I and IV alone was significant ($p = .011$, one-tailed test) in the high-suggestibility group.

The main findings may be summarized as follows:

1. Prestige increases body-sway suggestibility.
2. *Ss* high in suggestibility, unlike those low in suggestibility, showed a significant increase in body-sway under prestige.

## REFERENCES

EYSENCK, H. J. *Dimensions of personality*. London: Kegan Paul, 1947.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

(Received July 21, 1959)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 488–490

# SOCIAL CONFORMITY IN PERCEPTION OF THE AUTOKINETIC EFFECT[1]

NORMAN S. ENDLER[2]

*Division of Counseling, Pennsylvania State University*

SHERIF (1935) used the autokinetic (AK) technique[3] to study the development and maintenance of social norms, and concluded that social norms evolve in group situations and may subsequently manifest themselves when the subject (*S*) is alone. Since the *S*s use subjective scales in developing AK norms, the experimenter (*E*) cannot objectively control the norms.

Luchins (1954) modified Sherif's technique by varying the illumination surrounding the AK light. He and Paul (1950) have shown that the point of light becomes fairly stable over trials for an *S* after the surrounding illumination reaches a certain intensity. Thus, the amount of illumination needed to stabilize the point of light (for an *S*), the point of cessation (of AK movement), is a reliable measure of the threshold below which the *S* observes the AK movement. Three advantages of his method are: it introduces stimulus evidence, it is objective, and the social norms can be controlled by *E*.

In this study Luchins' technique was used to test two hypotheses: (*a*) the AK point of cessation can be raised or lowered by special conformity pressures; and (*b*) the socially induced changes are relatively permanent, i.e., true conformity rather than compliance occurs.[4]

## METHOD

The AK apparatus was a black light-proof wooden box, 12″ × 12″ × 14″, with a 4″-square aperture in the center of its front face. The box was divided by a vertical black cardboard with a 2.8 mm. center hole. A 40-w. bulb behind the partition (the AK stimulus) burned continuously. A 100-w. bulb connected to a variac, placed in front of the partition served as the variable objective field surrounding the AK light. Twenty grade school and high school students were randomly divided into two groups: "stopped moving" group (9 boys, 1 girl); and "always moving" group (7 boys, 3 girls). *S*s were tested under four different sequential conditions or treatments.

First, a pretreatment reference point was established. *S*, seated 12 ft. from the front of the AK box, was told to report everything he saw. After *S* first reported movement of the AK light, *E*, seated behind *S*, gradually increased the surrounding illumination from zero variac units until *S* reported cessation of the AK movement. The variac reading at this point, when converted to watts, was an index of the point of cessation. The illumination was then increased to a maximum of 119 w. (ascending series). For the descending series the illumination was decreased from 119 w. until *S* reported AK movement again. This point also served as an index of the point of cessation. Three ascending and three descending series were averaged separately[5].

One day after the pretest, *S* was tested in the presence of an accomplice. For the Stopped Moving group, the experimental aim was to lower the *S*'s point of cessation. Pretest procedures were modified as follows: The accomplice always responded before *S*. On *E*'s signal the accomplice began reporting cessation of movement at a level of illumination well below the point where *S* usually reported cessation and continued until maximum illumination was achieved. For the Always Moving group, the aim was to raise *S*'s point of cessation. The procedure was the same as for the Stopped Moving group except that the accomplice always reported AK movement irrespective of illumination.

Posttests I and II were the same for both groups and were exactly the same as the pretest. Posttest I occurred immediately after the experimental treatment. Posttest II occurred two weeks later.

The points of cessation were averaged,[6] producing one ascending and one descending threshold score (in watts) for each *S* for every treatment.

## RESULTS AND DISCUSSION

Means and standard deviations are presented in Table 1. Four analyses of variance of the two-factor type were computed, and showed in each case that differences among session means were significant ($p < .001$).[7]

[1] This paper is based on a master's thesis conducted at McGill University in 1954. The author is indebted to A. S. Luchins who directed this thesis, for his helpful comments and suggestions. A modified form of this paper was presented at the Illinois Psychological Association in April 1957 in Peoria, Illinois. The author also wishes to thank Donald H. Ford for his aid with the final draft.

[2] Now at York University, Toronto, Canada.

[3] The perception of movement of a stationary pinpoint of light, fixated in a completely dark room, is called the autokinetic effect. This effect can be induced in most *S*s and creates a situation which is new to most *S*s.

[4] For this study the empirical basis suggested by Festinger (1953) for distinguishing between true conformity and compliance was used. True conformity implies that changes due to social pressure are maintained in a subsequent nonsocial situation, and compliance implies that socially induced changes are not maintained in subsequent nonsocial situations.

[5] The ascending and descending series were analyzed separately because *S*s often complained about afterimages in the descending series.

[6] See Footnote 5.

[7] For the Always Moving group there were also sig-

TABLE 1
MEANS AND STANDARD DEVIATIONS (IN WATTS) FOR POINTS OF CESSATION OF AUTOKINETIC MOVEMENT

| Group | Series | Sessions | | | |
|---|---|---|---|---|---|
| | | Pretest | Accomplice Present | Immediate Posttest | Two-week Posttest |
| Stopped Moving | Ascending | | | | |
| | *M* | 12.76 | 7.82 | 8.30 | 7.47 |
| | *SD* | 1.18 | 0.65 | 1.05 | 1.15 |
| | Descending | | | | |
| | *M* | 11.24 | 7.52 | 8.30 | 6.04 |
| | *SD* | 2.05 | 0.60 | 1.05 | 2.39 |
| Always Moving | Ascending | | | | |
| | *M* | 11.37 | 15.31 | 15.69 | 16.22 |
| | *SD* | 1.59 | 3.21 | 3.74 | 4.14 |
| | Descending | | | | |
| | *M* | 10.48 | 16.46 | 15.69 | 14.56 |
| | *SD* | 1.36 | 4.29 | 3.74 | 3.27 |

Tukey's (1949) tests for comparing individual means in the analysis of variance revealed that the pretest means differed significantly ($p < .01$) from the treatment and posttest means. For the Stopped Moving group the treatment and posttest means were significantly larger ($p < .01$) than the pretest means. For the Always Moving group the opposite was true.

The results indicate that social pressure produced a significant change in the point of cessation of AK movement (a relatively stable norm). The point of cessation was regularly either raised or lowered depending on the nature of the social pressure and was in the direction of conformity to an accomplice's judgments. The new point of cessation was relatively stable, and manifested itself in nonsocial situations both immediately after, and two weeks after the *S*s were subjected to social pressure. Thus, apparently true conformity rather than public compliance occurred in the social pressure situation.

nificant differences among *S*s. Individual differences, however, were not of prime relevance to this study and therefore there were no further analyses of the differences among *S*s.

## REFERENCES


ENDLER, N. S. The influence of social factors on the autokinetic effect. Unpublished master's thesis, McGill University, 1954.

FESTINGER, L. An analysis of compliant behavior. In M. Sherif and M. O. Wilson (Eds.), *Group relations at the crossroads*. New York: Harper, 1953.

LUCHINS, A. S. The autokinetic effect and gradations of illumination of the visual field. *J. gen. Psychol.*, 1954, **50**, 29–37.

PAUL, I. H. The modified autokinetic technique. Unpublished master's thesis, McGill University, 1950.

SHERIF, M. A study of some social factors in perception. *Arch. Psychol., NY*, 1935, **27**, 187.

TUKEY, J. W. Comparing individual means in the analysis of variance. *Biometrics*, 1949, **5**, 99–114.


(Received July 28, 1959)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 491–494

# A FACTOR ANALYSIS OF THE WESLEY RIGIDITY INVENTORY:

## ITS RELATIONSHIP TO AGE AND NONVERBAL INTELLIGENCE

SHEILA M. CHOWN

*Medical Research Council, University of Liverpool*

THE Wesley Rigidity Inventory has been previously described in this journal by Wesley (1953). It consists of statements rated by five psychologists as indicating a high degree of rigidity in those who agree with them. Total score on the inventory has been found to correlate to a significant extent with other tests said to measure rigidity. For instance, it correlated with certain aspects of concept formation in a card sorting test (Wesley, 1953), with behavior in a group level of aspiration test (Zelen & Levitt, 1953), and with score on the California F Scale (Jones, 1954).

In the original form 50 items were used; later a 41-item version of the inventory was developed. Zelen and Levitt (1954) pointed out that a casual reading of the 41-item version suggested an overlap between a number of items. By inspection of the original statements, these authors chose 12 which they considered did not overlap, and proceeded to show that this form of the inventory was as reliable as the 41-item version, that the distribution of scores obtained with the two versions was similar, and that similar levels of correlation were obtained between the long and short forms and the California E Scale. However no factor analysis or item analysis appears to be available, so that there is no evidence to suggest that either form of the test is unidimensional. The object of the present analysis was to find out whether the inventory is unidimensional, and if not, to ascertain the relationship between age of subjects, nonverbal intelligence, and the separate dimensions.

### METHOD

*Procedure and Subjects.* The Wesley inventory (39 statements from the 41-item version) was given as part of a battery of 15 paper-and-pencil tests designed to investigate various aspects of rigidity in relation to chronological age. The battery, selected from reports of previous work on rigidity (Chown, 1959), also included Raven's Progressive Matrices (1938) and the Mill Hill Vocabulary (synonyms) Scale (Raven, 1943).

The battery was completed in group testing sessions by 200 male subjects aged between 20 and 82. Subjects were obtained from military hospitals, foreman courses, veterans clubs, Worker's Educational Association classes, and "over 60's" clubs.

Table 1 shows the age distribution of the sample. The distribution of Progressive Matrices scores was approximately normal, with a mean of 40.2 and a standard deviation of 9.1. On this measure of nonverbal intelligence the scores decreased slightly from 20 to 60 years and then more rapidly. The Vocabulary scores (mean 20.9 and standard deviation 6.0) tended to rise with age until 60 and then fall off.

*Analysis.* The true-false responses to the 39 items, chronological age, nonverbal intelligence, and vocabulary score for the 200 subjects were intercorrelated. Tetrachoric coefficients were used between items; biserial $r$'s for items with other tests; and product-moment $r$'s between age, nonverbal intelligence, and vocabulary. The distribution of all intercorrelations is shown in Table 2. The median correlation was 0.13.

TABLE 1

| Age | Number of Subjects |
|---|---|
| 20–29 | 54 |
| 30–39 | 50 |
| 40–49 | 33 |
| 50–59 | 39 |
| 60+ | 24 |

TABLE 2
DISTRIBUTION OF INTERCORRELATIONS IN ORIGINAL MATRIX

| Number of correlations | Size of correlations | $p$ |
|---|---|---|
| 456 | 0.00–0.13 | $>.05$ |
| 98 | 0.14–0.17 | .05–.01 |
| 81 | 0.18–0.22 | .01–.001 |
| 226 | 0.23–0.74 | $<.001$ |

The matrix was factor analyzed by the centroid method. McNemar's (1942) criterion was used to test for completeness of factor extraction, and factoring was stopped after the third factor had been obtained. Seven orthogonal rotations were carried out. The three factors together accounted for 24.1% of the variance.

### RESULTS

The items marked with an asterisk in Table 3 have been selected to form factor scores. Inspection of the loading patterns suggested that good representation of each factor could be achieved without serious loss of purity by adapting a general criterion loading of 0.24 for the inclusion or rejection of an item. (There is no generally accepted test of significance for a factor loading, but the figure chosen in this instance is close to that for twice the *SE* of a zero $r_{tet}$ for $N = 200$: viz., 0.22.)

TABLE 3
ROTATED FACTOR LOADINGS

| Item number | I | II | III |
|---|---|---|---|
| 1 | *29 | 04 | −01 |
| 2 | 04 | *48 | −04 |
| 3 | *48 | 00 | −17 |
| 4 | 11 | *28 | −19 |
| 5 | −02 | *49 | 11 |
| 6 | −09 | *42 | 22 |
| 7 | 10 | 14 | *52 |
| 8 | 09 | *43 | −08 |
| 9 | 20 | 16 | *26 |
| 10 | 14 | 03 | 17 |
| 11 | −29 | 30 | 20 |
| 12 | 50 | 28 | 16 |
| 13 | −17 | *40 | 22 |
| 14 | −38 | 26 | 36 |
| 15 | *39 | −10 | 18 |
| 16 | 13 | *36 | −04 |
| 17 | −24 | 23 | *34 |
| 18 | −17 | 20 | *43 |
| 19 | *56 | 16 | 12 |
| 20 | *63 | 16 | −24 |
| 21 | 07 | 39 | 30 |
| 22 | *59 | 14 | 10 |
| 23 | 19 | *66 | −10 |
| 24 | 42 | 25 | −08 |
| 25 | −04 | 20 | −12 |
| 26 | 36 | 25 | −05 |
| 27 | 12 | 19 | −08 |
| 28 | 16 | *42 | 24 |
| 29 | 04 | *30 | −12 |
| 30 | 10 | −10 | −17 |
| 31 | 15 | 14 | *33 |
| 32 | *34 | 16 | 10 |
| 33 | 40 | −17 | 39 |
| 34 | −12 | *53 | 00 |
| 35 | 28 | 40 | −39 |
| 36 | *41 | 18 | 18 |
| 37 | *60 | 06 | 18 |
| 38 | *36 | 14 | 20 |
| 39 | 10 | 06 | *33 |
| Raven | −04 | *−34 | −20 |
| Vocab. | 08 | −38 | 56 |
| Age | 14 | 12 | *63 |

Twenty-nine items were loaded above 0.24 on one of the factors and below 0.24 on the other two. There were 10 such items in Factor I, 12 in Factor II, and 7 in Factor III. The items are recorded in Tables 4, 5, and 6, where it can be seen that age loaded highly on Factor III and reversed Raven score or "unintelligence" on Factor II. Thirteen items were not counted in any factor, four because they were low on all, and nine (including Vocabulary score) because they were high on more than one.

The selected items, when summed with unit weighting, predict their respective factors 0.84, 0.90, and 0.68. Intercorrelations between the scores are as follows: I and II, 0.23; I and III, 0.17; II and III, 0.21. While these coefficients are all significant below the .01 level, their small size indicates that the three scores are not highly related to each other.

## DISCUSSION

Of the three factors, II is the easiest to identify. Both nonverbal intelligence and vocabulary have a high negative loading on it (see Table 3). Inventory items included in it consist of liking for slow steady methods, dealing with one thing at a time, sticking to own ways or beliefs, and statements about "people-in-general." It is not surprising to find that this type of rigidity is associated with low intelligence.

Factor III, with age having the highest loading, does not include many inventory items. The six which do load highly, however, present a consistent picture of liking for established routine.

Factor I, not associated with either intelligence or age, is made up of items dealing with persistence, methodical behavior, and liking for detail.

In view of the high correlation between age and nonverbal intelligence it had been thought that the analysis would produce one factor on which both age and intelligence would load highly, and perhaps another factor, not associated with either. The interesting thing about these results is that age and nonverbal intelligence do *not* load highly on the same factor. One group of items is related to chronological age and another to this measure of intelligence. The third group of items, as expected, is not related to either intelligence or age.

It is possible to examine the composition of Zelen and Levitt's short form of the inventory in the light of the present results. When describing their method of selection, they gave an example of two items which they thought were opposites and another example of two items which they thought so similar as to overlap. Their "opposite" items (present Numbers 12 and 31) do not load highly on the same factors. Moreover, their "overlapping" items (present Numbers 19 and 23) appear one in Factor I and the other in Factor II. These pairs are, it seems, not comparable after all, and presumably some information is lost by eliminating one of them. When their 12 final items are examined, it is found that 4 of these do not appear in the version used here, but 8 are common to the short and long versions. Five of the eight items are present in the factors, three loading highly on I, one on II, and one on III. Of the three remaining items, one loads highly on all three factors, one on I and II, and one on I and III. Thus it seems that consistent with the general contention of Zelen and Levitt the short form of the inventory approximates in factor make-up to the structure of the main inventory, though Factor II is underrepre-

TABLE 4
SCORED ITEMS FOR FACTOR I

| Item number | Loading | Item |
|---|---|---|
| 20 | 0.63 | I am usually able to keep at a job longer than most people. |
| 37 | 0.60 | I am (not) inclined to go from one activity to another without continuing with any one for too long a time. |
| 22 | 0.59 | I always finish tasks I start, even if they are not very important. |
| 19 | 0.56 | I am a methodical person in whatever I do. |
| 3 | 0.48 | I prefer work that requires a great deal of attention to detail. |
| 36 | 0.41 | I usually dislike to set aside a task that I have undertaken until it is finished. |
| 15 | 0.39 | My interests (do not) tend to change quickly. |
| 38 | 0.36 | I prefer to do things according to a routine which I plan myself. |
| 32 | 0.34 | I believe that promptness is a very important personality characteristic. |
| 1 | 0.29 | I am often the last one to give up trying to do a thing. |

TABLE 5
SCORED ITEMS FOR FACTOR II

| Item number | Loading | Item |
|---|---|---|
| 23 | 0.66 | People who go about their work methodically are almost always the most successful. |
| 34 | 0.53 | It is the slow steady worker who usually accomplishes the most in the end. |
| 5 | 0.49 | I prefer doing one thing at a time to keeping several projects going. |
| 2 | 0.48 | There is usually only one best way to solve most problems. |
| 8 | 0.43 | I usually maintain my own opinions even though many other people may have a different point of view. |
| 6 | 0.42 | I dislike to change my plans in the midst of an undertaking. |
| 28 | 0.42 | I have never done anything dangerous for the thrill of it. |
| 13 | 0.40 | I would not like the kind of work which involves a large number of different activities. |
| 16 | 0.36 | I usually find that my own way of attacking a problem is best, even though it does not always seem to work in the beginning. |
| Raven | −0.34 | Nonverbal intelligence |
| 29 | 0.30 | It is always a good thing to be frank. |
| 4 | 0.28 | I often become so wrapped up in something I am doing that I find it difficult to turn my attention to other matters. |

TABLE 6
SCORED ITEMS FOR FACTOR III

| Item number | Loading | Item |
|---|---|---|
| Age | 0.63 | Chronological age |
| 7 | 0.52 | I would (not) like a position which requires frequent changes from one task to another. |
| 18 | 0.43 | I (do not) like a great deal of variety in my work. |
| 17 | 0.34 | I dislike having to learn new ways of doing things. |
| 39 | 0.33 | I always put on and take off my clothes in the same order. |
| 31 | 0.33 | I (have not) taken a good many courses on the spur of the moment. |
| 9 | 0.26 | I find it easy to stick to a certain schedule once I have started on it. |

sented. Like the long version, it is not unidimensional, but consists of three factors and certain items loading highly on more than one of these.

The original correlation between age and total score on the Wesley was 0.27, and between the Matrices and the Wesley, −0.29. It is now obvious that these correlations arise from responses to different items in the questionnaire. The present results agree fairly well with conventional stereotypes of rigid behavior: people of low intelligence acting in a rigid manner because they only see certain aspects of a situation, the elderly being unwilling to change long established patterns of behavior, and some people of all ages displaying an orderly detail-loving nature. Factor scores should reveal more information about each individual than did the total score which, though not unidimensional, has been shown previously to yield useful data. The factor scores provide, possibly for the first time, a means of measuring the

extent to which different types of rigidity operate in an individual. It is not to be expected that a given piece of behavior will always stem from only one of these types of rigidity, but that in a given person at least these three factors will be combined in causing behavior, and that knowledge of the extent of the three types in any individual could be used, for example, in training procedures given to that person.

A factor analysis of the full battery of rigidity tests given to this sample will shortly be carried out, with these Wesley factor scores included in the matrix. In this way it should be possible to follow-up the three Wesley factors and link them to the different types of rigidity included in the battery. Later it is hoped to use the new Wesley subscores in industrial and experimental studies of learning.

SUMMARY

The Wesley Rigidity Inventory was completed by 200 subjects ranging in age from 20 to 81. The responses to 39 inventory items were intercorrelated and the resulting matrix factor analyzed by the centroid method. Age and scores on Raven's Progressive Matrices 1938 and Mill Hill Vocabulary (synonyms) were included in the matrix. Three orthogonal factors were found in the inventory, with age loading 0.63 on one factor and the Raven score −0.34 on another. It was found that the short form of the Wesley inventory devised by Zelen and Levitt also contains items from the three factors. It is proposed to use these new Wesley subscores in future work with the inventory.

REFERENCES

CHOWN, SHEILA M. Rigidity—a flexible concept. *Psychol. Bull.*, 1959, **56**, 195–223.

JONES, M. B. Aspects of the autonomous personality. *US Naval Sch. Aviat. Med. res. Rep.*, 1954, Proj. No. NM 001058.25.16.

MCNEMAR, Q. On the number of factors. *Psychometrika*, 1942, **7**, 9–18.

RAVEN, J. C. *Progressive Matrices (1938)*. London: Lewis, 1938.

RAVEN, J. C. *The Mill Hill Vocabulary Scale*. London: Lewis, 1943.

WESLEY, ELIZABETH. Perseverative behavior in a concept-formation task as a function of manifest anxiety and rigidity. *J. abnorm. soc. Psychol.*, 1953, **48**, 129–134.

ZELEN, S. L., & LEVITT, E. A group level of aspiration technique as a measure of personality rigidity. *Proc. Iowa Acad. Sci.*, 1953, **60**, 569–572.

ZELEN, S. L., & LEVITT, E. Notes on the Wesley Rigidity Scale: The development of a short form. *J. abnorm. soc. Psychol.*, 1954, **49**, 472–473.

(Received August 3, 1959)

*Journal of Abnormal and Social Psychology*
1960, Vol. 61, No. 3, 495–496

# INTROVERSION-EXTRAVERSION AND KINESTHETIC AFTEREFFECTS

ALLAN RECHTSCHAFFEN[1] AND LAWRENCE J. BOOKBINDER[2]
*University of Chicago Medical School* *Northwestern University*

EYSENCK (1955) hypothesized that extraverts develop greater neural satiation than introverts and offered evidence which he interpreted as supporting this hypothesis. He used dual criteria of psychiatric diagnosis and scores on the Rhathymia (R) scale (Guilford, 1942) to define the introversion-extraversion dimension and amount of kinesthetic aftereffect (KAE) to measure neural satiation. His results showed a tendency for the extraverts to develop greater KAE. A more recent study (Rechtschaffen, 1958) disclosed no significant differences in amount of visual aftereffect (VAE) developed by introverted and extraverted college students, where R score alone was used to define the personality dimension. One possible explanation for the negative result in the latter study derives from the particular sense modality used in measuring aftereffect, i.e., a relationship between aftereffect and introversion-extraversion might have been obtained had KAE been used rather than VAE. The present study was designed to explore this possibility by comparing the KAEs of introverted and extraverted college students.

## METHOD

*Subjects.* The R scale was administered to a large introductory psychology class. Students with the lowest R scores were requested to serve as the introvert subjects, those with the highest R scores as the extravert subjects. Few students refused to serve; thus, the experimental sample represents the vast majority of those asked to participate. The mean R score of the extravert group ($N = 37$) was 53.9; the $SD$ was 4.9, and the range 48 to 69. The mean R score of the introvert group ($N = 37$) was 26.7; the $SD$ was 4.9, the range 16 to 34. There were 17 males and 20 females in the extravert group and 21 males and 16 females in the introvert group.

*Apparatus.* The apparatus used to measure KAE was similar to the one used by Eysenck (1955). It consisted of a comparison scale, a test object, and two stimulus objects; all were blocks of polished hardwood. The comparison scale was 30 in. long, 2 in. deep, and tapered from 4 in. to 0.5 in. The stimulus and test objects were each 6 in. long and 2 in. deep. One stimulus object was 2.5 in. wide and the other 0.5 in. wide, the test object was 1.5 in. wide.

*Procedure.* Subjects were tested individually and were blindfolded during the entire experiment. They were given five practice trials each, a trial consisting of four adjustments of the comparison scale, with a 1-minute rest between trials. The procedure for a practice trial was as follows: the subject stood in front of a table on which were mounted the comparison scale and the test object. The thumb and middle finger of the preferred hand were positioned in a movable rider on the comparison scale, so as to grasp the sides of the scale. The corresponding fingers of the other hand grasped the sides of the test object. Subjects were instructed to move the rider along the comparison scale until they judged that the width of the comparison scale was the same as the width of the test object. Four such adjustments were made in rapid succession. Starting position of the rider and distance of the comparison scale from the subject were systematically varied on successive adjustments to reduce position and movement cues.

After a 1-minute rest following the fifth practice trial, the subject was given the first preinspection trial, which was identical to a practice trial. Immediately following the preinspection trial, the comparison scale and test object were replaced by the 2.5-in. and 0.5-in. stimulus objects, respectively. Then for 40 seconds the subject vigorously rubbed the sides of the 0.5-in. stimulus object with the thumb and middle finger of his nonpreferred hand, and at the same time he rubbed the 2.5-in. stimulus object with the thumb and middle finger of his preferred hand. This inspection period was immediately followed by a postinspection trial, which also followed the same procedure as the practice trials. This procedure was intended to produce a "double satiation" effect: i.e., theoretically, stimulation of the test object hand by the smaller stimulus object should make the test object appear wider following satiation, and stimulation of the comparison scale hand by the wider stimulus object should make the comparison scale seem smaller than it did prior to inspection. Both changes in perception should work together to produce wider adjustments on the comparison scale following inspection.

Three more measures of KAE, each measure consisting of a preinspection trial, inspection, and a postinspection trial, were made during the experimental session. A 3-minute rest separated each of the four measures. Amount of KAE was measured as the absolute difference between average preinspection and postinspection adjustments of the comparison scale for each trial.

## RESULTS

The results do not support the hypothesis that extraverts develop more KAE than introverts. Analysis of variance showed no significant differ-

[1] A major portion of this research was done while the senior author was at the Chicago Branch of the Veterans Administration Neuropsychiatric Research Laboratory, Veterans Administration West Side Hospital.

[2] Now at Veterans Administration Hospital, Ann Arbor, Michigan.

ence between personality groups on amount of KAE over the four trials. Neither was there a significant difference between introverts and extraverts on practice or preinspection trials, thus eliminating the possibility that group differences in constant error confound comparisons of amount of KAE.

For both personality groups, amount of KAE on the first trial was considerably larger than on the last three; the appropriate $F$ ratio showed the decrease of KAE over successive trials to be significant at the .001 level. That this decrease results, at least in part, from incomplete dissipation of KAE during the 3-minute rest intervals is suggested by the failure of preinspection trials to return to original practice levels ($F$ ratio for preinspection trials was significant at the .001 level). Thus, the results of the first KAE trial become crucial, since subsequent trials may become confounded with residues of undissipated KAE; large first trial KAEs may preclude large second trial KAEs because of increased preinspection adjustments, combined with relatively asymptotic postinspection adjustments. Therefore, introverts and extraverts were also compared with respect to first trial KAEs alone. The mean first trial KAEs for introverts and extraverts were .32 and .29 in. per adjustment, respectively. The $t$ value for the difference was .66, a nonsignificant value.

Neither do the present data support Eysenck's hypothesis that extraverts develop more persistent aftereffects than introverts, since the interaction term for trials × personality group was not significant at the .05 level for either preinspection or postinspection trials.

A sex difference in KAE was found. Males had a mean first trial KAE of .36 in. as compared to .25 in. for females; the $t$ value for the difference was 2.12, significant beyond the .05 level. In view of the above difference and the unequal number of males and females in the two personality groups, separate comparisons of introverts' and extraverts' KAEs were made for each sex. Mean KAE for male introverts was .40 in. as compared with .33 in. for male extraverts; the $t$ value was .83, not significant. Mean KAE for female introverts was .27 in. as compared with .22 in. for female extraverts; the $t$ value was .73, also not significant. Thus, although there were slightly different numbers of males and females in the two personality groups and a significant sex difference in KAE, these factors did not seem to alter the major finding that extraverts showed no evidence of more KAE than introverts.

## SUMMARY

Eysenck's hypothesis that extraverts show greater figural aftereffects than introverts was investigated by comparing the kinesthetic aftereffects of 37 subjects with high scores on Guilford's Rhathymia scale (extraverts) with those of 37 subjects with low scores on the scale (introverts). The two personality groups did not differ significantly in amount of aftereffect. Several differences remain between the present study and Eysenck's earlier one, but the generality of the hypothesis that extraverts develop greater aftereffects than introverts is seriously questioned.

## REFERENCES

EYSENCK, H. J. Cortical inhibition, figural aftereffect, and theory of personality. *J. abnorm. soc. Psychol.*, 1955, **51**, 94–106.

GUILFORD, J. P. *An Inventory of Factors STDCR.* Beverly Hills, Calif.: Sheridan Supply, 1942.

RECHTSCHAFFEN, A. Neural satiation, reactive inhibition, and introversion-extraversion. *J. abnorm. soc. Psychol.*, 1958, **57**, 283–291.

(Received August 3, 1959)

*A complete library of psychology in one volume—*

# CLASSICS IN PSYCHOLOGY

**Edited by Thorne Shipley**

In a text of over 1400 pages, are the works of 35 leading psychologists of Europe, the United States, and the Soviet Union, spanning a period of one hundred and fifty years.

*Some of the chapters:*

**PHILIPPE PINEL:**
Treatise on Insanity

**JEAN MARTIN CHARCOT:**
Diseases of the Nervous System

**WILHELM WUNDT:**
Sensory Perception

**ERNST MACH:**
The Relation of the Physical to the Psychical

**WILLIAM JAMES:**
Principles of Psychology

**EMIL KRAEPELIN:**
Clinical Psychiatry

**E. B. TITCHENER:**
Structural Psychology

**BENJAMIN RUSH:**
The Diseases of the Mind

**MANFRED SAKEL:**
Shock Treatment of Schizophrenia

**EUGEN BLEULER:**
Dementia Praecox

**ALFRED ADLER:**
Individual Psychology

**I. P. PAVLOV:**
Conditioned Reflexes

**HERMANN RORSCHACH:**
Psychodiagnostics

**AUGUST AICHHORN:**
Wayward Youth

**G. STANLEY HALL:**
Adolescence

**JEAN PIAGET:**
The Language and Thought of the Child

**KURT KOFFKA:**
Perception

**ISAAC RAY:**
Jurisprudence of Insanity

**WILLIAM MCDOUGALL:**
Social Psychology

**JOSEF BREUER & SIGMUND FREUD:**
Studies on Hysteria

**C. G. JUNG:**
Analytical Psychology

$20.00

*You can expedite shipment by enclosing remittance.*

**PHILOSOPHICAL LIBRARY, Publishers**

15 East 40th Street New York 16, N.Y.

# *Announcement*

**For the benefit of our new subscribers and others who may wish to complete their files of PSYCHIATRY, we are offering a special discount on large orders of back issues. This discount, of 20 percent from the regular prices given below, will apply only to orders of twenty or more issues. Our regular prices:**

**Single issues, 1938 through 1953, $2.00 ($2.20 foreign)**

**Single issues, 1954 on, exclusive of current year, $2.50 ($2.70 foreign)**

**Bound volumes (4 issues), 1938 through 1953, $12.50 ($13.30 foreign)**

**Bound volumes (4 issues), 1954 on, $14.50 ($15.30 foreign)**

***Issues published within the calendar year are available by subscription only.***

ORDER FROM: **PSYCHIATRY,**

**1610 New Hampshire Ave., N.W.**
**Washington 9, D. C.**

***All orders are prepaid; make checks payable to***
**PSYCHIATRY, A PUBLICATION**

*Just published—*

## THE PSYCHOLOGY OF CRIME

*By David Abrahamsen, M.D.*

A detailed and provocative analysis of the criminal mind by an internationally known psychoanalyst and author. Basing his findings on intensive research and many case histories, Dr. Abrahamsen deals with the emotional and social derangement which carries the criminal into action. There is a full chapter each on the juvenile delinquent, the sex offender, and the murderer.

It is Dr. Abrahamsen's firm conviction that the solution to the crime problem lies in rehabilitation and prevention rather than retaliation and punishment. To this end he surveys present-day provisions for handling offenders and suggests how they could be changed to keep up with modern psychiatric knowledge.

*The Psychology of Crime* is an important—and challenging—new book.

Price: *$6.00*

**CONTENTS**

Social Pathology and Crime

The Formula of Criminal Behavior

Family Tension

Juvenile Delinquency

Psychosomatic Disorders and Crime

The Offender and the Emotionally Disturbed Nonoffender

The Classification of Criminals

The Acute and the Chronic Offender

The Sex Offender

The Personality of the Murderer

Psychiatric-Psychological Examination of the Offender

Criminal Law and Psychiatry

Rebuilding and Rehabilitating the Offender

The Prevention of Crime

## THE ANATOMY OF PSYCHOTHERAPY

Systems of Communication and Expectation. *By Henry L. Lennard and Arnold Bernstein. With Erdman B. Palmore and Helen C. Hendin, M.D.* The authors demonstrate in detail how the concepts and approach of sociology can yield new understanding of and insight into the problems faced by psychotherapy. Based on over 500 tape-recorded sessions from eight individual therapies, this book will be of great interest to those who work in sociology or psychology. *$6.00.*

## PATHS OF LONELINESS

*By Margaret Mary Wood.* In this book, a leading sociologist analyzes today's lonely people and the causes for their aloneness. With compassion and understanding, she shows how the processes of individualism, competition, segregation, and stratification wear down the ability of people to establish lasting contact with each other. Columbia Paperback No. 3: *$1.50.* Also available in clothbound edition: *$3.75.*

**Columbia University Press**

**2960 Broadway New York 27**

*Second printing . . .*

# RESEARCH IN PSYCHOTHERAPY

Proceedings of an interdisciplinary conference,
Washington, D. C., April 9–12, 1958
ELI J. RUBINSTEIN and MORRIS B. PARLOFF,
*Editors*

Continued demand for this provocative report, out of print for the past six months, necessitated reprinting; copies are again available.

**Price $3.00**

*Order from:*

**AMERICAN PSYCHOLOGICAL ASSOCIATION**
**Department RP 2**
**1333 Sixteenth Street, N.W.**
**Washington 6, D. C.**

BOOKS

LEISURE IN AMERICA:
A Social Inquiry

By Max Kaplan, *Boston University*. This book presents a detailed and systematic analysis of the patterns, the functions, and the problems of leisure in American society. It is not, however, another contribution to that already large body of literature which tells how leisure time *should* be spent or criticizes how it is spent. Rather it is a work of objective scholarship and investigation which deals with the basic questions that must be answered before really meaningful suggestions or criticisms can be made. Such questions as: What is leisure? How can various forms of leisure be classified? What are the currently observable trends in the use of leisure time? What is the relationship of various types of leisure activity to important variables in American society? In attempting to answer these and other important questions, the author presents many new and significant insights into leisure as a cultural phenomenon. *316 pages. $7.50.*

THE PSYCHOLOGY
OF CHARACTER DEVELOPMENT

By Robert J. Havighurst, *University of Texas*, and Ruth G. Taba, *University of Chicago; with* Ruth Cavins, James Anderson, *and* Dorothy McKone, *all of the University of Chicago*. Formulates a new, systematic theory of character development based on modern psychodynamic principles. The theory was developed from—and tested by—rigorously conducted research studies in which the concepts of "classical" psychology were combined with valid, sophisticated research methodology. The explanation of character development given in the book is comprehensive and takes into account both motivational and environmental factors. The book may also serve as a source of reliable, accurate data for assessing personality and character structure. In addition, the authors cite factual evidence pointing toward the development of a true science of ethics based on the observable facts of human behavior and their reactions to various kinds of situations. *[illegible] pages. Prob. $5.00.*

*Send for examination copies*

JOHN WILEY & SONS, Inc.

440 Park Avenue South New York 16

"The best in tests. . . . . "

Could you use the new, 1960 edition of this well-known test for brain injury and perceptual-motor coordination, widely used in clinical practice and in research since 1945, with new norms based on large clinical and control groups?

Graham-Kendall

# MEMORY FOR DESIGNS TEST

NEW EDITION!!

Then order Tester's Set, consisting of durable plates and 1960 manual ($8.50), or just the 1960 manual ($3.50) if you already have plates.

•

Obtain further information by writing

PSYCHOLOGICAL TEST SPECIALISTS
Box 1441
Missoula, Montana

"The best in tests. . . . . "

NEW EDITION!!